Signature Page

RAUDING THEORY AND THE TOEFL SUBTESTS

A Dissertation

Submitted to

the Temple University Graduate Board

in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF EDUCATION

By Gabriel A. Lee

January, 2002

ABSTRACT

Rauding Theory and the TOEFL Sub-Tests
Gabriel A. Lee
Doctor of Education
Temple University, 2002
Dr. James Dean Brown

Rauding theory states that two factors, accuracy and rate, explain nearly all the variance in the general reading ability of American schoolchildren (Carver,1999). Using Structural Equation Modeling (SEM) and regression, this dissertation investigated to what extent the rauding model could be applied to 92 Japanese university aged young adults studying English as a second language. The TOEFL sub-tests defined the latent factor efficiency level (EL). Two vocabulary tests defined the latent factor accuracy level (AL), two speed of comprehension tests defined the latent factor rate level (RL). A modified version of the Woodcock Non-Word Recognition test and a phonological awareness test defined the latent variable pronunciation level (PL). Results indicated moderate support for the rauding model. In the SEM analysis, a non-significant chi-square and a CFI of .98 indicated good model fit. 62% of the individual differences in the TOEFL sub-tests (EL) could be explained by regarding AL and RL as correlated factors. PL, on the other hand, did not correlate with EL at all. In addition, RL correlated with AL at .66 ($p < .05$). The only aberration with Carver's rauding model was that the correlation between PL and RL was non-significant at $p < .05$. A regression analysis reconfirmed the results of the SEM analysis. An

indicant of EL, defined as (RL + AL) / 2, predicted 62% of the variance in the TOEFL sub-tests. Finally, a measure of cognitive speed significantly correlated with RL at $p < .05$.

These results have four main implications. First, the rauding model needs to be investigated further in the context of second language acquisition. Second, language proficiency, or performance on tests of language ability, can be defined as efficiency level. Third, performance differences in language tests purporting to measure reading and listening proficiency may be attributable to the latent factor RL, not PL. Finally, instruction focusing on improving RL should be incorporated into the Japanese national curriculum.

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

CHAPTER 1

CONTEXT

The idea that an important relationship exists between speed and proficiency in a foreign language is not new. Indeed, there is an almost instinctive logic behind it. Experience tells us that the more skillful we become at doing something, the less we have to pay attention to what we are doing and the faster we are able to do it. This idea plays a fundamental role in the two major theories concerning skill acquisition in language learning as well as most accounts of skill acquisition in psychology (Anderson, 1985; Anderson, Lebiere, & Lovett, 1998; Cohen, McClelland & Dunbar, 1990; LaBerge & Samuels, 1974; Logan, 1979; McLaughlin, 1990; Palmeri, 1997; Posner & Snyder, 1975; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977).

Tests of language comprehension, be they tests of reading or listening, traditionally purport to measure only one factor: accuracy of comprehension. That is to say they claim to measure *how accurately* one has comprehended a passage. Recent research in first language (L1) reading comprehension, however, challenges this assumption. Carver (1992b, 1992c) has provided evidence suggesting that either two factors: rate of comprehension and accuracy of comprehension, or one factor: efficiency of comprehension, explain virtually all the variance in standardized tests of reading comprehension with native speaker schoolchildren. Other L1 researchers and theorists also have suggested that the concepts of efficiency and rate are important to the measurement of reading

comprehension (Jackson & McClelland, 1979; Stanoivch, 1986). Indeed, Perfetti (1985) has argued that one cannot define reading ability without considering both speed and comprehension.

In the field of second language acquisition, however, relatively little attention has been focused on factors of rate or efficiency of comprehension. A digital search for the terms “reading rate” and “speed” on the CD ROM version of *TESOL Quarterly* (1967 to 1996) shows that thirty years ago more researchers seemed to be interested in this question than today. From 1967 to 1973, six articles appeared in the *TESOL Quarterly* which either used reading rate as a research variable or discussed the importance of reading rate as a marker of ability (see Esky, 1970; MacNamara, 1970; Norris, 1970; Oller, 1972; Pierce, 1973; Plaister, 1968; Saville-Troike, 1973). In the following twenty-two years (1974 to 1996), however, only two other articles appeared, both of which used reading speed as a research variable (Macha, 1979; Tomiyama, 1980). *TESOL Quarterly*, read by language instructors and researchers alike, is one of the major journals in the study of second language acquisition and instruction. The results of this search suggest that among the readership of the *TESOL Quarterly* minimal attention has been paid to the relationship between reading rate and second language (L2) ability in the last twenty five years.

Considering the salience of rate of comprehension and efficiency of comprehension as theoretical constructs, it seems important to find out how much these factors influence scores on tests of general English ability such as

the TOEFL. In a nation such as Japan, where tests of English ability are used to make decisions affecting millions of students'admission to universities and high schools, it would seem particularly important that educators understand the relationship between test performance and these factors.

Rauding Theory

Rauding theory grew out of the research of Ronald Carver (1977, 1982, 1983, 1984, 1985a, 1985b, 1987a, 1987b, 1987c, 1990a, 1990b, 1992a, 1992b, 1992c, 1992d, 1993, 1997, 1999) and is an attempt to explain the causes of high and low reading achievement for native speakers of English. The term "rauding" was derived from a combination of two words, reading and auding; reading usually means understanding language via printed words and auding understanding language via spoken words. Rauding is therefore defined as the process of understanding complete thoughts while either reading or auding. One of the main tenents of rauding theory is that there is relatively little difference between the process involved in comprehending relatively easy text when reading and comprehending the same text when it is read aloud at the typical reading rate of the individual.

During reading, rauding is similar to what is often called "simple reading," or "normal reading." Individuals who are rauding are said to be comprehending all or almost all of the complete thoughts in the text as they are being read.

Furthermore, they are reading at their normal reading rate, which is relatively constant and limited by their cognitive speed.

There are three laws which form the foundation of rauding theory and hence the rauding model discussed below (Carver, 1987). The first law states that individuals attempt to comprehend thoughts in a passage at a constant rate, unless they are pressured by situation specific factors to alter that rate. An individual could be pressured to go slower, for instance, if the material was conceptually very difficult or had many unknown words. The same individual could be pressured to go faster if that person only had a short amount of time and was looking for specific information like a date or a name. Barring such situation specific factors, however, rauding theory assumes that all individuals have a constant rate of comprehension that is called the rauding rate of the individual.

Law two of rauding theory states that the efficiency of passage comprehension depends on the accuracy and rate of passage comprehension. Equation (1) expresses this relationship as:

$$E = AR \tag{1}$$

According to equation (1), if an individual reads the sentences in a text at the rate of 20 standard sentences (one standard sentence equals one hundred character spaces) per minute (R = 20 Spm) and understood 80% of them (A =

.80), then this individual's efficiency of comprehension would be 16 standard sentences per minute (E = AR = (.80)(20) = 16).

Law three of rauding theory states that individuals' most efficient rates of comprehension are their rauding rates. When reading rate is equal to the individual's rauding rate, then E will be at its highest. If rate does not equal the rauding rate of the individual, then E will be lower.

A great deal of empirical evidence exists that supports these three laws (e.g. Carver, 1982, 1983, 1984, 1985a, 1985b). Because they serve as the foundation for the rauding model, which is reviewed in detail below, further discussion will focus on the rauding model only.

In the rauding model for reading achievement there are six theoretical constructs that explain student performance on tests of language ability. These are Efficiency Level (EL), Rate Level (RL), Accuracy Level (AL), Verbal Level (VL), Pronunciation Level (PL), and Cognitive Speed Level (CS). The model defines reading achievement as rauding efficiency level (EL). According to Carver (1999), EL has two proximal causes: rauding rate level (RL) and rauding accuracy level (AL). RL is an individual's typical reading rate, while AL refers to the most difficult text an individual can accurately read when it is read at the individual's normal reading rate. Simply put, AL is a measure of an individual's reading vocabulary knowledge, while RL represents the fastest speed that the individual can comfortably comprehend the text. The definition of EL in rauding theory is (AL + RL) / 2. This means that an individual with a high reading level

(AL = 8th grade level) but a low reading rate (RL = 3rd grade level) would only be able to perform at the fifth grade level (EL = 5.5) on standard tests of reading comprehension.

With native speakers who are still beginning or intermediate readers, AL is theorized to have two proximal causes: verbal level (VL) and pronunciation level (PL). VL is measured by how many words an individual can understand auditorily and refers to the most difficult text an individual can understand when it is read aloud at the individual's normal reading rate. In the causal model, AL is hypothesized to equal the average of VL and PL, that is AL = (VL + PL) / 2. For advanced readers (PL > 8th grade level), however, VL = AL (Carver, 1999).

The other proximal cause of EL is RL. RL refers to an individual's normal reading rate when there is neither time pressure to go fast nor any need to have a special knowledge of the material after reading. In Carver's (1999) causal model, RL has two proximal causes with beginning and intermediate readers: PL and cognitive speed (CS) and is defined as (PL + CS) / 2. However, with advanced readers RL is defined as simply RL = CS. This means that once an individual has mastered the sound system of English, PL ceases to be a factor influencing RL.

In the causal model, it is clear that PL is the most important factor for educators working with native speaker children to consider because it influences both AL and RL. PL is defined in Carver's (1999) causal model as the average of an individual's vocabulary knowledge (AL) and decoding knowledge (DK).

Decoding knowledge is similar to cipher knowledge and represents the ability to correctly pronounce pseudo words, whereas PL represents the ability to pronounce actual words. This can be written as PL = (AL + DK) / 2. Carver (1999) has argued that an individual's spelling knowledge (SP) almost perfectly correlates with PL. This means that in rauding terminology, pronunciation level can also be defined as PL = SP.

The relationships between these six theoretical constructs in the causal model can be seen in below in Figure 1. Currently only research with native speakers of English has been used to support this model. Note that all causal relationships in Carver's causal model are based on logic, not statistical inference based on experimental research. Correlation does not imply causation; the claim of causation can only be ascertained through experimental research where variables are directly manipulated.

Rauding theory was originally devised to explain reading achievement for native speakers of English. When it is applied to non-native speakers of English, however, some theoretical issues arise. First is the definition of AL as (PL + VL) / 2 with beginning or intermediate readers. With native speakers of English who are beginning or intermediate readers, it is safe to assume that they understand more through listening than reading. However, with non-native speakers this may not be a valid assumption. Learners of English in Japan, for instance, may be able to understand a word in its printed form but not its spoken form. This means that VL cannot be a proximal cause of AL. Furthermore, since VL is not a

proximal cause of AL, it is unlikely that PL is a proximal cause of AL either. For PL to be a proximal cause of AL, it is assumed that an individual can understand the spoken form of the word first; therefore, when it is silently pronounced during reading, it is understood. This problem probably does not exist with advanced level non-native speakers of English. For them, it is possible that VL and AL are equal. Indeed, it is logical to assume that most advanced learners of English learn the written and oral form of a word at the same time.

A second problem is the definition of RL as (CS + PL) / 2 for beginning or intermediate readers. Most native speakers can be said to be fluent in their L1. Because of this, only CS and PL influence RL. This makes sense because a native speaker individual with poor pronunciation ability will have difficulty reading and comprehending quickly (note that in rauding theory, pronunciation ability refers to an individual's ability to read real words, while decoding ability refers to an individual's ability to read non-words). The same cannot be said,

Table 1. Six Theoretical Constructs, Their Symbols, Their Corresponding Traditional Concepts, Their Equations, and Their Commonly Used Names

Symbol	Theoretical Construct	Similar Traditional Concepts	Equation	Common Name
EL	Rauding efficiency level	Reading achievement, general reading ability, or ability to read efficiently.	(AL + RL) / 2	Efficiency level
AL	Rauding accuracy level	Reading level, or most difficult text that can be accurately comprehended during reading.	(VL + PL) / 2	Accuracy level

RL	Rauding rate level	Normal reading rate	(PL + CS) / 2	Rate level
VL	Verbal knowledge level	General knowledge, or the most difficult text that can be accurately comprehended during listening.		Verbal level
PL	Pronunciation knowledge level	Decoding ability, or the number of words that can be accurately identified	(AL + PL) / 2 SP	Pronunciation level
CS	Cognitive speed level	Rate of naming letters or numbers, thinking speed.		Cognitive speed level.

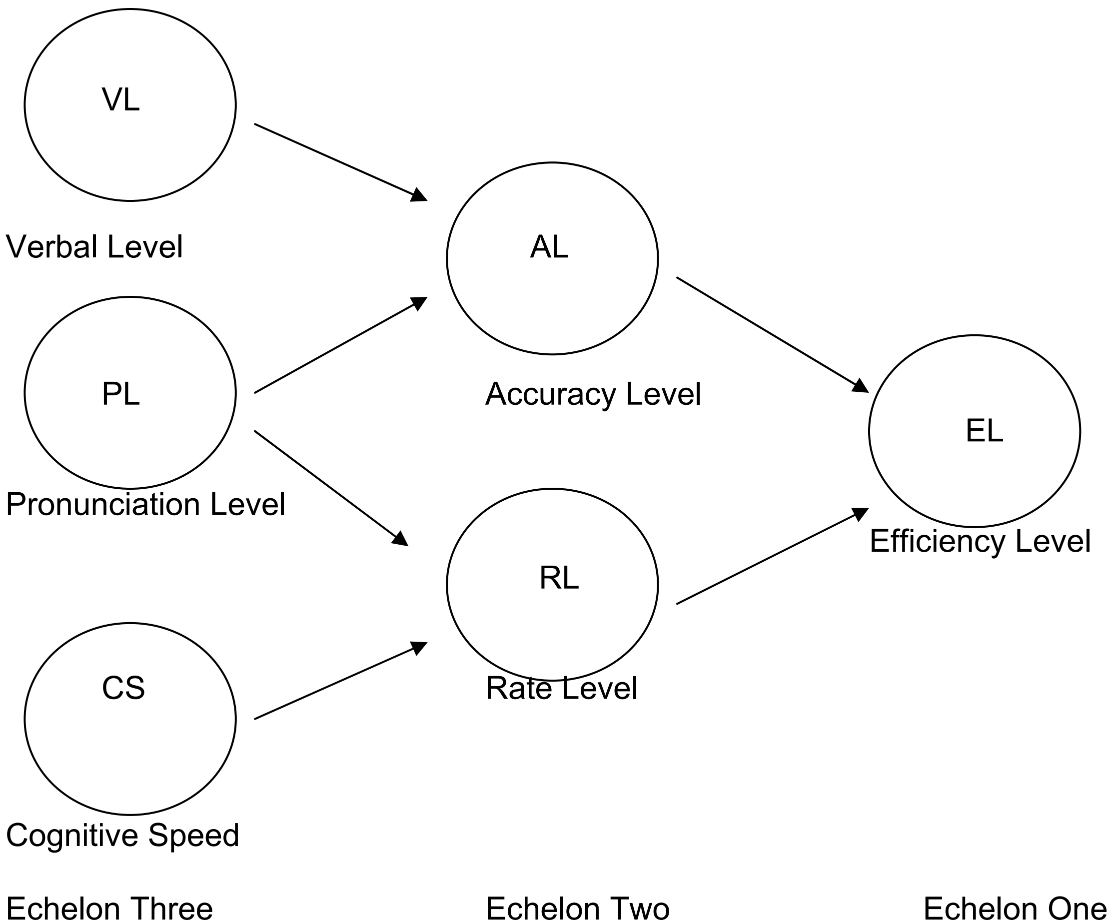

Figure 1. The Causal Model for Reading Achievement in Rauding Theory

however, for non-native speakers of English who may know the written form very well but not the oral form of the word. Such readers might be able to read quickly but unable to pronounce the words well. In addition, most non-native speakers are not fluent in their L2. This means that RL is less influenced by CS and may be more influenced by other factors such as vocabulary, amount of language exposure or type of language exposure (mostly hard or mostly easy materials) to name a few.

Another problem is that with non-native speakers it is unclear whether PL is as important as it is for native speakers. Indeed, as stated earlier, it is doubtful that PL directly influences either AL or RL. For non-native speakers who have learned the written form of the word before the oral form, it may be that PL = (VL + DK) / 2 rather than (AL + DK) / 2. The role of PL with non-native speakers, therefore, needs to be further examined.

Finally, how accurate are the formulas that make up rauding theory when they are applied to second language learners? Does EL = (AL + RL) / 2, or could EL = (AL + (RL/2)) / 2? Indeed, there is some evidence from prior research reviewed below that AL explains more of the variance in EL than does RL. This question too needs to be examined further.

Statement of Purpose

The purpose of this study is to investigate whether or not the causal model for reading achievement proposed by Carver (1999) is applicable to the test performance of non-native speakers of English. It is hypothesized that Carver's causal model is valid up to echelon two but not at echelon three. What role do VL, PL, and CS play in non-native speaker test performance? How do they relate to the variables in echelon two and one? What modifications will need to be made to the causal model for non-native speakers? These are a few of the

issues it will address. Below are the nine hypotheses to be investigated in this study.

Hypotheses

H1 Rate Level (RL) will significantly correlate ($p < 0.05$) with General English Ability, or EL, as measured by the total TOEFL score.

H2 Accuracy Level (AL) will significantly correlate ($p < 0.05$) with General English Ability.

H3 Efficiency Level will significantly correlate ($p < 0.05$) with General English Ability as measured by the TOEFL with a corrected correlation coefficient of .75 or higher.

H4 RL will correlate higher with the listening section (T-List) of the TOEFL than with either the structure (T-Struc) or reading comprehension (T-Reading) sections.

H5 RL will correlate significantly ($p < 0.05$) with cognitive speed (CS).

H6 RL will correlate, but not significantly. ($p > 0.05$), with pronunciation level (PL).

H7 RL and AL will be highly correlated. ($p < 0.05$)

H8 PL will equal AL

H9 Hypothesized Structural Model shown in Figure 2

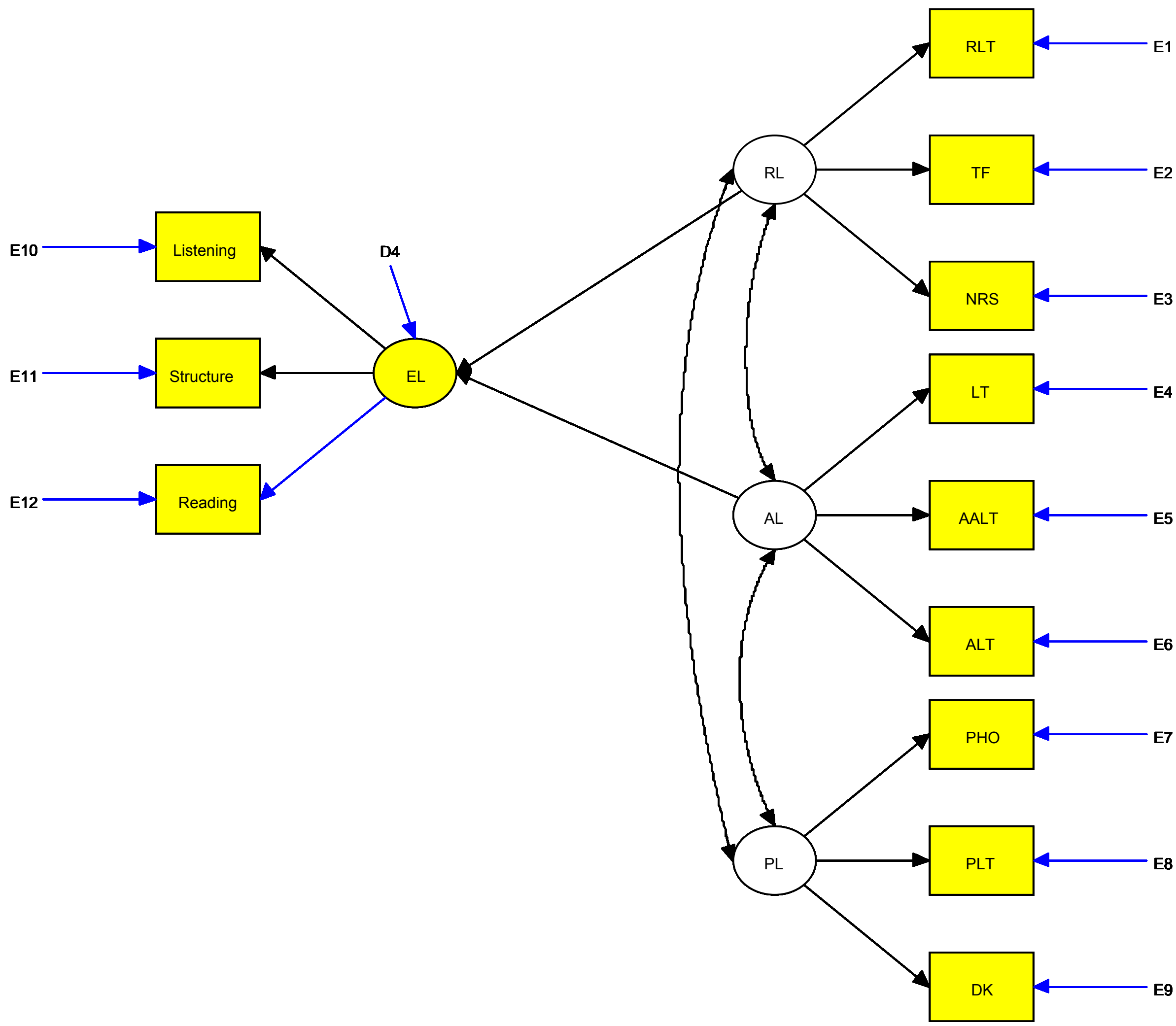

Figure 2. Hypothesized Structural Model

EL = Efficiency Level, *RL* = Rate Level, *AL* = Accuracy Level, *PL* = Pronunciation Level, *Listening* = TOEFL Listening, *Structure* = TOEFL Structure, *Reading* = TOEFL Reading, *RLT* = Rate Level Test, *TF* = True-False Rate of Comprehension Test, *NRS* = Natural Reading Speed Test, *LT* = Nation's Levels Test, *AALT* = Auding Accuracy Level Test, *ALT* = Accuracy Level Test, *Pho* = Phonological Processing Skill Test, *PLT* = Pronunciation Level Test, *DK* = Non-Word Pronunciation Level Test

CHAPTER 2

LITERATURE REVIEW

The causal model shown graphically in Figure 1 can be depicted in precise mathematical equations. This is advantageous because it allows the model to be easily falsified. The following section will examine theory and data relevant to the relationships posited by the causal model for both native speakers of English and non-native speakers of English.

Rauding Theory and Native Speakers of English

Efficiency Level or EL = (AL + RL) / 2

In rauding theory, efficiency level is similar in concept to reading achievement. Carver defines it as "the highest level of text difficulty (DL) at which an individual can accurately comprehend (accuracy (A) > .64) when the individual is allowed to read the text at an average rate that is equal to the difficulty level of the text when both are in grade equivalency (GE) units, that is, RL = DL" (Carver, 1999, p. 47). The decision to use .64 as the criterion for accuracy of comprehension comes from data collected from native speaker school children performance on standardized tests of reading ability. The data show that when DL matches AL (DL = AL), or when relative difficulty is zero (DL -

AL = 0) that 64% of the passage will be comprehended if read once at the individual's RL, that is, normal reading speed (Carver, 1990b).

Therefore, EL can be measured with most standardized tests of reading achievement. However, there is one caveat. The more a standardized test relies on inferential questions or is made up of reading relatively hard materials, the less it will be an accurate measure of EL. That is, the more the test requires novel problem solving such as usually required by reading a difficult text and answering inferential questions, the more that fluid intelligence (GF) becomes a causal factor.

One goal of the EL construct is to provide a yardstick by which students at various ability levels can be measured in grade equivalency (GE) units so that after one year of schooling, students should increase one GE unit. Since reading achievement tests that rely heavily on the ability to make inferences show less yearly gain due to the quality of schooling, Carver (1999) has argued that they are not appropriate for judging the adequacy or quality of one year of education. Therefore, such tests are not appropriate measures of EL in the causal model.

Another concept similar to EL is general reading ability. In 1954, Holmes discussed general reading ability as being composed of "speed" and "power," which translates directly into efficiency level because EL is made up of AL (accuracy or power) and RL (rate or speed). More recently, Perfetti (1989) devoted an entire article arguing for the existence of a general reading ability. Noting that most arguments contrary to the existence of a general reading ability

are based upon the role of general and specific knowledge (e.g., schema theory), he argued that general knowledge cannot be identified with general reading ability, although specific knowledge can sometimes be vital for reading comprehension.

Perfetti's ideas also fit the causal model; general reading ability, or EL, cannot be directly affected by general knowledge, also called verbal knowledge (VL), because it is also affected by pronunciation level (PL) and cognitive speed (CS) (Carver, 1999). Indeed, several studies have shown that with native speakers, phonemic awareness is a more powerful predictor of literacy acquisition than generalized measures of intelligence (Bradley & Brant, 1983; Stanovich, Cunningham, & Freeman, 1984; Share, Jorm, Maclean, & Matthews, 1984; Tunmer & Nesdale, 1985). Moreover, Perfetti argued that general reading ability is not greatly affected by the ability to make inferences or elaborations, which also fits well into the causal model since EL is not greatly affected by fluid intelligence either.

In conclusion, L1 reading theory supports the concept of general reading ability and can be used to support the existence of EL as a theoretical construct. EL, however, is precisely defined and therefore empirically falsifiable. According to the causal model, an indicant of EL should be able to be separated into two components, AL and RL. Therefore, standardized tests of reading comprehension which (a) have a time limit and (b) do not require novel problem

solving, should be able to be broken down into either one factor, EL, or two factors, AL and RL.

Carver (1992c) investigated in a series of four studies to what degree standardized tests measure the constructs EL, AL, and RL. Subjects ranged from grade three to college. Scores on the Iowa Test of Basic Skills (ITBS), Nelson-Denny Reading Comprehension Test (NDRT), and the Degrees of Reading Power test (DRP) were analyzed because they were considered representative of widely used standardized tests in American schools.

Scores on the ITBS were reported according to the test's four sections: (a) Vocabulary test (ITBS-Vocab); (b) Reading Comprehension test (ITBS-Comp); (c) Language Skills test (ITBS-Lang); and (d) Math Skills test (ITBS-Math). ITBS-Vocab was hypothesized to be an indicant of AL and ITBS-Comp an indicant of EL. Neither ITBS-Lang nor ITBS-Math was hypothesized to be an indicant of either EL, AL, or RL. Likewise, scores on the NDRT were reported according to the test's three sections: (a) Vocabulary (ND-Vocab); (b) Rate score (ND-Rate); and (c) Comprehension (ND-Comp). ND-Vocab was hypothesized to be an indicant of AL, ND-Rate an indicant of RL, and ND-Comp an indicant of EL. The DRP was hypothesized to be an indicant of AL because it is an untimed reading test.

In addition to the ITBS, NDRT, and DRP, three other tests were also administered. The Rauding Accuracy Level Test (ALT), the Rauding Rate Level Test (RLT), and the Rauding Efficiency Level Test (RELT). The ALT is a

vocabulary test and was hypothesized to be a measure of AL. The RLT is a timed cloze test in which the correct answer is always obvious. This test was hypothesized to be a measure of RL. The RELT contains 100 word passages that vary in difficulty from Grade 1 to Grade 18. As passages increase in difficulty, the amount of time allocated for reading the passages is reduced. The score on the test is determined when an individual can no longer answer the comprehension questions after each reading with greater than 75% accuracy. This test was hypothesized to be a measure of EL. An additional variable was created by averaging the GE-score of the ALT and RLT to create an estimate of EL based on the equation EL=(AL + RL) / 2.

A principal components analysis was used in each of the four studies reported. Each study presented two analyses. In Analysis 1 the number of eigen values greater than one was used to determine the number of factors. If Analysis 1 resulted in a single-factor fit, then in Analysis 2 a two-factor fit was forced, and if Analysis 1 resulted in a two-factor fit, then Analysis 2 forced a single-factor fit. When there were two factors, an oblique rotation (direct oblimin) was used because prior data collected in 1972 by Carver and Darby suggested that the factors were correlated. The best measure of each factor was empirically determined by the highest loading on that factor. That is, if the highest loading was on the RLT, a measure of RL, then the factor would be named a rate level factor.

In study one, scores from 354 students ranging from grade three to grade eight were analyzed. Variables were the ALT, RLT, ITBS, as well as a measure of efficiency based on the average of the ALT and RLT. Published reliability estimates were reported for each test, but actual reliability estimates were not. Analysis 1 resulted in a single-factor that explained 71% of the variance. The variable that loaded highest on this factor was the measure of efficiency based on the average of the RLT and ALT (.90). The second highest loadings were on the ITBS-Comp (.87) and ITBS-Lang (.87) respectively. This factor was therefore named an efficiency factor and the hypothesis that ITBS-Comp was a measure of EL was supported.

In Analysis 2 a two-factor fit was forced. The highest loading on factor one was with the ITBS-Vocab (.90), considered to be a measure of AL. The second highest loading was with ITBS-Lang (.86), followed by the ALT (.85) and ITBS-Comp (.85). This factor was therefore named an accuracy factor. The variable that loaded highest on factor two was the RLT (.97), followed by ITBS-Comp (.67) and ITBS-Lang (.67). This factor was therefore named a rate factor. In conclusion, the hypothesis that scores on the ITBS were influenced by either one factor (EL) or two factors (AL and RL) was supported. Furthermore, the hypothesis that EL is comprised of two subfactors, AL and RL, was supported.

In study two, data from a subgroup of the students from study one who had also taken the Rauding Efficiency Level Test (RELT) were reported ($N = 56$). Again, published reliability estimates were reported for each test, but actual

reliability estimates were not. Analysis 1 resulted in a single factor that explained 75.7% of the variance. As found before, the two highest loadings were on the efficiency variable created from the average of the ALT and RLT (.95) and the ITBS-Comp (.91). The factor was therefore defined as an efficiency factor and the hypothesis that ITBS-Comp is a measure of EL was again supported. The RELT, however, loaded at .80 which was lower than expected, suggesting that the RELT was not as good a measure of EL as the others. Reliability estimates were not reported, however, so it may be possible that low reliability with the RELT was responsible for its lower factor loading. Analysis 2 resulted in two factors with results almost identical to Analysis 2 in study one. The highest loading on the first factor was the ITBS-Vocab (.94) and the highest loading for factor two was on the RLT (.93). In conclusion, study two supported the contention that the ITBS was measuring either one factor (EL) or two factors (AL and RL).

In study two, data from a subgroup of the students from study one who had taken the Degrees of Reading Power (DRP) were reported (*N* = 102). These students ranged from grade seven to grade eight. Once again, published reliability estimates were reported for each test, but actual reliability estimates were not. Results were very similar to study one and two. In Analysis 1 a single factor accounted for 81.1 % of the variance. The highest loading was again on the measure of EL derived from the average of the ALT and RLT (.94), followed by the DRP (.92) and the ITBS-Comp (.91). When a two-factor fit was forced, the

highest loading on factor one was once again the ITBS-Vocab (.93), followed by the DRP (.92) and the ALT (.90). This finding supports this factor's being called an accuracy level factor rather than a vocabulary factor or a knowledge factor for two reasons. First, the DRP is an untimed, criterion-referenced measure of reading comprehension requiring students to answer questions about increasingly difficult passages. Second, it is not a vocabulary test (Carver, 1992). Finally, as predicted, the highest loading for factor two was on the RLT (.99) followed by the ITBS-Comp (.75) and the DRP (.72). The fact that the DRP loaded so highly on the rate factor was unexpected because it is an untimed test. However, limits on concentration as well as the nature of administering a test at a school may make an untimed test impossible, which may be why the DRP loaded almost as highly as the ITBS-Comp on the rate factor.

Study four examined the performance of 64 colleges students on the Nelson-Denny Reading Test (NDRT). The variables examined were the ND-Vocab, ND-Comp, ND-Rate, RLT, ALT, and a measure of efficiency derived from the average of the RLT and ALT scores. Reliability estimates were reported for all measures: ND-Vocab, .84; ND-Comp, .68; ND-Rate, .86; ALT, .91; RLT, .83; and Efficiency, .88. Analysis 1 resulted in two factors that correlated .41 and explained 74.1% of the variance. The first factor had its highest loadings on ND-Vocab (.92) and the ALT (.90). Since both of these measures were hypothesized to be measures of AL, factor one was called an accuracy factor. Factor two was interpreted as a rate level factor because its highest loading was with the RLT

(.91), followed by ND-Rate (.65), both of which were hypothesized to be measures of RL. In Analysis 2 a single-factor fit was forced upon the data. It explained 55.6% of the variance and was identified as an efficiency factor because the highest loading was on the efficiency measure (.87) that was derived from the average of the RLT and ALT. This was followed by the ND-Comp (.79), which was also hypothesized to be a measure of EL.

In conclusion, Carver's (1992) four studies provide strong correlational evidence that standardized tests of reading comprehension are measuring either one factor, EL, or two factors, AL and RL. Furthermore, they support the idea that EL is comprised of two sub-factors, AL and RL. Carver's analyses would have been strengthened if reliability estimates had been available for studies one, two and three, but given the nature of standardized testing at the elementary and secondary school level, published reliability estimates may be the best that can be done.

Efficiency Level in Listening or EL = (VL + CS) / 2

The construct of Rauding Efficiency Level (EL) should not limited to simply a more precise definition of reading achievement. Indeed, the term rauding was created from two words, reading and auding, in order to focus upon the similarities between reading comprehension and listening comprehension. Remember, one of the assumptions of rauding theory is that the comprehension

of sentences, whether they are (a) being read as they are looked at in printed text, or (b) being auded as they are read aloud by someone else, is fundamentally the same process. This relationship can be expressed mathematically with the following equations.

Given:

$$EL = (AL + RL) / 2 \tag{2}$$

$$AL = VL \text{ with advanced readers} \tag{3}$$

$$CS = RL \text{ with advanced readers} \tag{4}$$

Then

$$EL = (VL + CS) / 2 \tag{5}$$

In 1990, Cunningham, Stanovich, and Wilson administered 22 reading related measures to 76 introductory psychology students. Among these measures were several measures relevant to equations (3) and (4) above. The Nelson Denny Reading Test (NDRT) provided a measure of AL from the vocabulary section (ND-Vocab). The NDRT was also administered aurally, providing a measure of VL (ND-Listening). The best measure of RL was a test called Easy Word RT, which measured the latency of vocal reaction time (RT) for naming easy words after they appeared on a monitor. The best indicant of CS came measuring latencies for naming 25 letters of the alphabet.

Split half reliabilities were reported as: .63 for ND-Listening (VL), .96 for ND-Vocab (AL); and .93 for Letter Naming RT (CS). No reliability estimate was reported for Easy Word RT. The corrected correlation between VL and AL was .73, which can be taken as support for equation (2) above, even though the causal model would predict a correlation closer to 1.00. The correlation between RL and CS was .54, which probably would have been higher if a reliability estimate for RL had been available and the correlation coefficient had been corrected for attenuation. Nevertheless, this is close to the correlation reported by Carver (1988) between the Speed of Thinking Test (a measure of CS) and the Rate Level Test (a measure of RL), which ranged between .59 and .53 for 129 college students. This provides limited evidence for equation (3), and perhaps suggests that RL is influenced by factors other than CS, although the causal model currently asserts that CS is the only causal factor influencing RL for advanced readers.

In conclusion, Cunningham, Stanovich and Wilson's (1990) research provide support for equation (3) and limited support for equation (4), therefore supporting the contention that efficiency of listening and reading comprehension are essentially the same process. That is, equation (2) can be transformed into equation (5). If this transformation is valid, we would expect to find evidence that EL in auding and EL in reading are the same construct for adults.

Carver (1982) set out to find such evidence when he tested the hypothesis that the optimal rate for comprehending thoughts while reading also

equals the optimal rate for comprehending thoughts while auding. In this study, 102 college students were presented passages at four difficulty levels: Grade 5, Grade 8, Grade 11, and Grade 14. The passages were presented both visually and auditorily. There were twenty passages at each grade difficulty. Subjects read ten of these passages and auded ten. The ten reading passages were presented at exactly the same rates as the ten auding passages. Rate was controlled in the reading mode by using motion picture film to present each passage and in the auding mode by using time-compressed speech. Rates varying from 83 Wpm (Wpm stands for standard length words, each of which is six characters long) to 500 Wpm were used. There were three measures of comprehension, two objective and one subjective. No matter which of the three was used, the efficiency of comprehension (E) was highest at rates close to 300 Wpm. This means that if the rate of passage presentation in either the auding or reading mode was higher or lower than 300 Wpm, the efficiency of comprehension decreased.

The results of Carver's (1982) study provide support for equation (2), (3), (4) and (5). By showing that the most efficient rate for reading comprehension is the same as the most efficient rate for listening comprehension, it provides direct support for the existence of a CS factor in auding and supports the relationships stated in equations (2), (3), (4) and (5).

Other evidence supporting a cognitive speed factor in auding comes from research with learning-disabled (LD) children. Cermak (1983), for example,

found that young LD children process information more slowly no matter if they are reading or listening. In addition, Kagan (1982) found that a reading disabled group was much slower than a control group in responding to aural stimulus. In conclusion, CS seems to be a factor that influences the efficiency of comprehension in both reading and listening and the fundamental process behind the efficiency of reading and listening comprehension seems to be the same.

Rauding Theory and Non-Native Speakers of English

So far all of the evidence pertaining to rauding theory and the causal model have been from studies where the native language of the subjects was English. Indeed, rauding theory and the causal model were derived from native speaker data and all of the research used to support it originates from research with native speakers. Nevertheless, several studies in EFL or ESL contexts provide evidence supporting aspects of Carver’s causal model and will be reviewed below. With the exception of one (Hirai, 1999), however, rauding theory and the causal model were not directly mentioned. Therefore, in order to show clear connections between these studies and Carver’s causal model, a reanalysis of data based on the principles of rauding theory was performed when necessary.

In 1972, John Oller collected data on 11 reading related variables with 50 learners of English at the University of California at Los Angeles (UCLA). Five of these variables were measures taken with eye movement photography (EMP). The other six were sub-tests of the UCLA ESL Placement Exam (ESLPE).

Oller examined the correlations between these variables and concluded that there was a predictive relationship between EMP measures of reading and other integrative language skills. He also noted that the correlation between the total score on the ESLPE and its vocabulary sub-test was .76, while the correlation between the total score on the ESLPE and its dictation sub-test was

.90. Based on these findings, Oller suggested that tests of language proficiency should attempt to measure one unified factor rather than a series of discreet factors.

Oller's findings and conclusions are easily explained by Carver's causal model. First, the causal model assumes a relationship between measures of reading and other integrative language skills because they are both measures of efficiency level, or EL. Second, the total score on the ESLPE and the score on a dictation test are both measures of EL, while a vocabulary test should be a measure of AL. Therefore, it is logical that two measures of EL would correlate higher than a measure of EL and AL.

Because there were several measures relevant to the relationships posited by Carver's (1999) causal model, a SEM reanalysis of Oller's data seemed possible. Reading Rate (Rate) was hypothesized to be a measure of Rate Level (RL). The vocabulary section of the UCLA ESL Placement Exam (ESLPE-Vocab) was hypothesized to be a measure of Accuracy Level (AL). The interview section (ESLPE-Inter) and dictation section (ESLPE-Dic) were hypothesized to be measures of efficiency level (EL). No reliability estimates were reported for any of the variables.

Using the correlation matrix reported by Oller (1972) a structural equation model was hypothesized in which the indicator variables ESLPE-Vocab (AL) and Rate (RL) define the latent variable EL, which in turn predicts the indicator variables ESLPE-Dic (EL) and EL-Inter (EL). Initially, a causal relationship was

hypothesized between AL and RL. This was different from the model outlined by Carver (1999), which posited a correlational relationship between AL and RL. For non-native speakers of English, however, it seemed logical that learners who encountered a large number of unfamiliar words would be forced to read slower. Likewise, learners who had large vocabularies would encounter few unknown words and would thus be able to read faster. Finally, it seemed illogical for a fast reading rate to have a causal effect on vocabulary size; therefore a directional relationship was hypothesized from AL to RL. After further assessment, however, it seemed unlikely that AL could be a primary cause of RL if the test used to measure RL used simple words. Thus, a correlational relationship was hypothesized.

This model is shown below in Figure 3. Maximum likelihood estimation was employed for estimation. The independence model testing the hypothesis that all variables are uncorrelated was easily rejectable, X^2 (6,N = 50) = 126.892, p < .05. The hypothesized model was tested next. Support for the hypothesized model was found in terms of a non-significant chi-square value and high CFI value, X^2(1,N = 50) = 0.146, p > .05, CFI = 1.0. The Bentler-Bonett Normed Fit Index (BBFI) and Non-Normed fit indexes (BBNFI) were also high, .999 and 1.042, indicating good model fit.

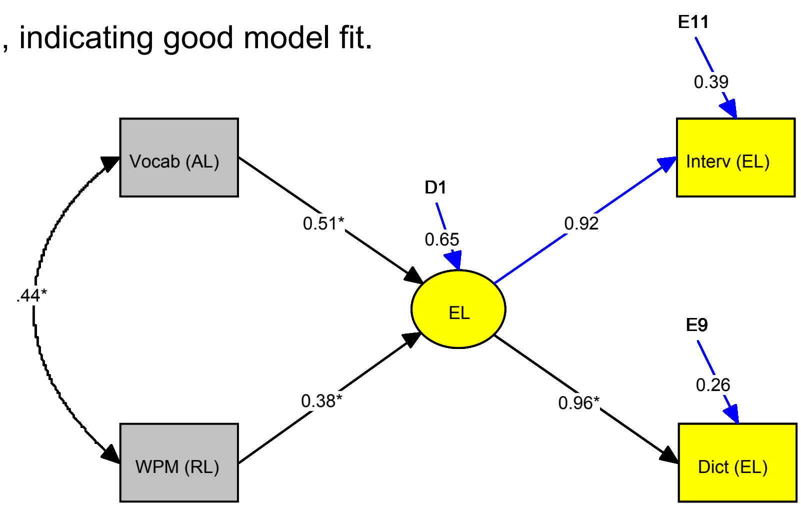

Figure 3. Structural Equation Model for Reanalysis of Oller's (1972) Data

The results indicate a good model fit. The use of a correlation matrix to analyze the data, however, may result in imprecise values for the parameter estimates, particularly for estimation of their standard errors (Schumacker & Lomax, 1996). Therefore, these results need to be interpreted with some caution.

These results provide moderate correlational evidence that EL (as measured by an oral interview and dictation) can be separated into two factors, AL and RL. The standardized coefficient for AL is 0.51 and for RL is 0.38. This means that AL explains slightly more variance than RL for this sample of the population. This is different from Carver's model, which predicts the two should have equal coefficients. Finally, this data suggest that the definition of EL needs to be modified to encompass more than simply the efficiency of reading or listening comprehension. Indeed, if EL can predict performance on an oral interview as well as dictation, it may be that it represents efficiency in general language ability rather than simply in reading and listening.

In conclusion, Oller's data and conclusions, as well as the reanalysis of his data, seem congruent with Caver's causal model. The fact that none of the instruments used in his study had reliability estimates, however, makes it difficult

to know how accurate the correlations were, and thus weakens any conclusions that can be made from his study or the reanalysis of his data.

Buck (1992) reviewed the results of a multi-trait multi-method study in which two theoretical traits (listening and reading comprehension) were tested across three methods (picture recognition, gap-filling, and self-rating). This study is directly relevant to the causal model because according to Carver's causal model, the fundamental process underlying reading and listening comprehension when AL = VL are the same.

The participants were 258 Japanese college and junior college students in and around Osaka, Japan. There were six tests: three listening tests and three reading tests, which were divided into the three methodologies named above. The first test was called Listening Picture Recognition (LR). It asked students to look at a paper and select a picture that corresponds to the description presented aurally. The K-R21 reliability estimate for this test was .83. The second test was Reading Picture Recognition (PR). This test was similar to PL except students read the descriptions and then circled the pictures. The K-R21 reliability estimate for this test was .82. The third test was called Listening Gap-Filling (GL). This test was a dictation-cloze type of test in which students were given a text with 24 words deleted. They listened to the text once and then they had five minutes to complete the test. The split half reliability estimate for this test was .83. The fourth test was called Reading Gap Filling (GR). A passage in which parts of 24 words were deleted was used. Students had 20 minutes to

complete this task. The split half reliability estimate for this test was also .83. The fifth test was called Listening Self Rating (SL). In this test students were asked to rate their listening ability on a Likert scale for eight different listening situations. The alpha reliability estimate for this test was .92. The final test was called Reading Self Rating (SR). This was similar to SL except the students read the passages instead of auding them. The alpha reliability estimate for this test was .88.

In his conclusion, Buck reported that “the [Buck/Ross] study failed to find clear evidence for the existence of a separate listening trait in second language English students” (1992, p. 332). This suggests that listening and reading comprehension are the same factor and provides moderate support for the validity of the causal model with ESL learners.

Structural Equation Modeling was used to reanalyze the data from the Buck/Ross study to test this hypothesis. Based upon the conclusions in Buck (1992), a four factor model was hypothesized: three factors representing the three different methodologies and one factor representing EL for both listening and reading. The initial run resulted in a Comparative Fit index of .99 and a chi-square of .04. In addition, a condition code indicated that EQS held the disturbance of the third factor (EL) at zero rather than permit it to become negative. These results were judged to be significantly good enough to pursue a post-hoc Lagrange multiplier test. The Lagrange multiplier test indicated that a correlation between the error terms E1 and E3 was significant.

Although many statisticians caution against correlating error terms in SEM, the components of error terms can be separable with multi-trait multi-method (MTMM) studies or panel models (Joreskog, 1993). Correlated error variances in MTMM studies are probably the result of non-random method variance (Maruyama,1997) and thus correlating error terms in this type of reanalysis should be helpful. Another way to say this is that the parts of PL and GL that were not common to the method Picture Identification or Gap Filling (i.e., the construct of listening efficency) covaried. This seemed plausible; thus the correlation between the error terms was added and the model tested again. Results are shown below in Figure 4. A non-significant chi-square of .733 and a Comparative Fit index of 1.0 indicated a good model fit.

In conclusion, the Buck/Ross data, Buck's (1992) reanalysis of that data, and the SEM reanalysis all provide support for the contention that the fundamental process underlying reading and listening comprehension is the same. This provides moderate support for rauding theory and the causal model, which hold that the construct EL is the same for both listening and reading. The use of the Lagrange Multiplier test and resultant correlated error terms, however, weakened the generalizability of the results beyond this particular data sample. Were the study to be replicated, it would be important to adjust the measurement instruments so that there were no correlated error terms.

Buck (1992) reported the results of another study in which he tried to demonstrate that listening and reading comprehension were separate traits.

Subjects were 353 university and junior college Japanese students from the Osaka area. This population as well as the design were very similar to the 1988 Buck/Ross study described earlier except the testing methodologies were slightly different. In this new study, four different methodologies were used: short answer, gap-filling, translation, and multiple-choice. Like in the 1988 Buck/Ross study, listening and reading were tested with three methodologies.

The first test was Listening: Short Answer Comprehension Questions (SL). In this test students listened to various passages and either gave a short answer in English or Japanese or filled in a grid. The listening passages were on topics such as an explanation of a school timetable, a telephone request to do a favor, and an invitation to a party. Reliability, as measured by Cronbach's Alpha, was .84.

The second test was Reading: Short Answer Comprehension Questions (SR). Students gave short answers to questions in English or Japanese. The reading passages included a timetable, a film guide, and an advertisement. The test/retest reliability coefficient was .80.

The third test was Listening: Gap-Filling (GL). The task asked students to fill in blanks on a Japanese summary of a story that they had listened to in English. Texts consisted to two different narratives told by two different speakers. Reliability, measured by Cronbach's Alpha, was .80.

The fourth test was Reading: Gap-Filling (GR). This was almost the same as GL except students read passages instead of listening to them. The test/retest reliability coefficient was .83.

The fifth test was Listening: Translation (TL). In this test students listened one long talk divided into several short sections. After each short section, they wrote down what they had heard in Japanese. The topic of the talk was a Canadian describing Canada for the benefit of a prospective Japanese tourist. Reliability, measured by Cronbach's Alpha, was .83.

The sixth test was Reading: Translation (TR). Students translated short passages they had read in English into Japanese. The texts were taken from tourist brochures describing two hotels in England. The test/retest reliability coefficient was .90.

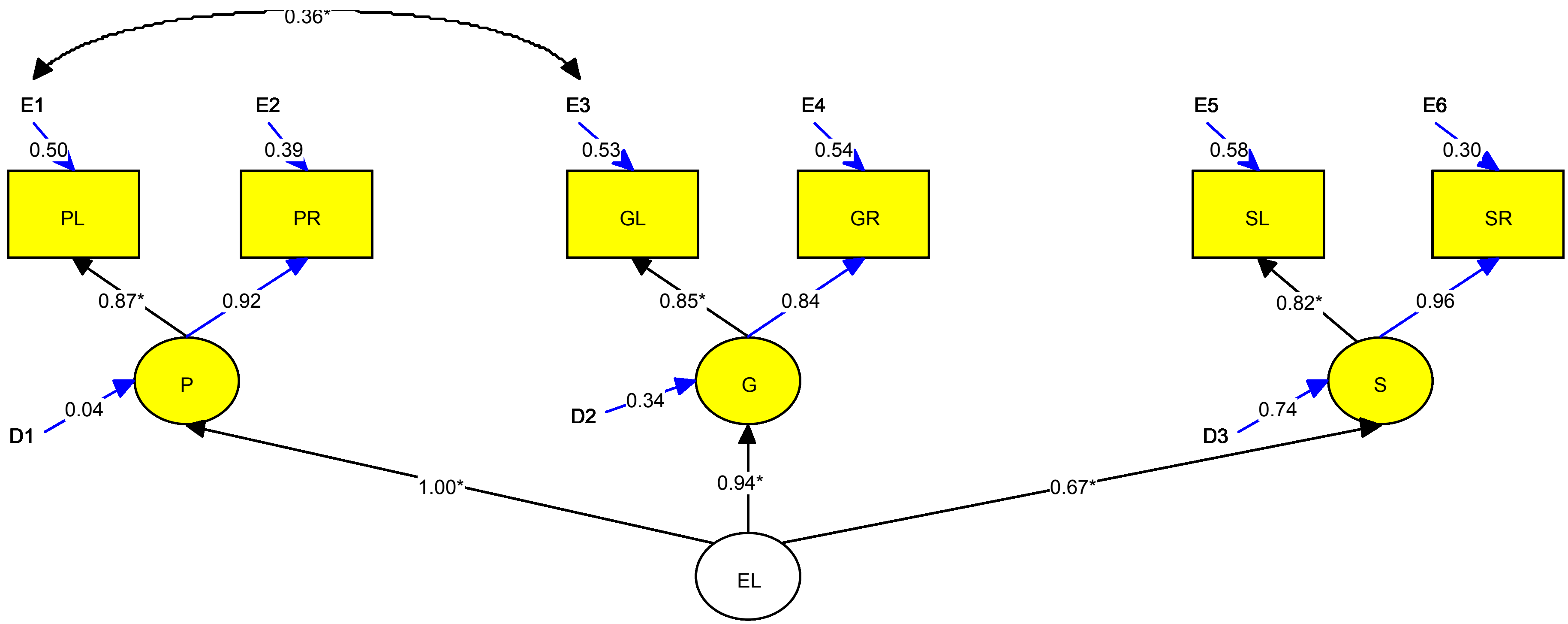

Figure 4. A Reanalysis of the 1988 Study by Buck/Ross Showing No Difference Between Listening and Reading Comprehension

The seventh test was Listening: Multiple-choice (ML). Students listened to 12 passages and answered 44 questions, each with 4 options. Subject matter was reported to be “various” and mainly written by the researcher. Reliability, measured by Cronbach’s Alpha, was .71.

The final test was Reading: Multiple-choice (MR). There were 16 passages, 33 items, each with four options. Texts were expository, taken from encyclopedias and other reference books. The test/retest reliability coefficient was .60.

Buck (1992) created a correlation matrix from the variables listed above and then analyzed it by means of the Campbell and Fiske multitrait-mulitmethod convergent discrimination validation technique. In this technique a series of tests (methods) are used to measure two or more traits. The data is put together in a correlation matrix and examined for convergent and discriminant validity. Using this technique, Buck found a “very clear trait effect” (1992, p. 345) among the eight tests. He reported that the listening tests met all of the Campbell and Fiske criteria for construct validity and the reading tests met them in 40 of 48 comparisons. Use of the Campbel and Fiske technique is questionable, however, as it assumes that the traits are uncorrelated, which they are not. An analysis of variance confirmed the Campbel and Fiske conclusion; there was significant trait effect, but a nonsignificant method effect. Buck (1992) concluded by saying that listening and reading comprehension cannot be considered one trait, but neither

can they be considered two entirely different traits due to the high correlation between them.

These conclusions seemed incomplete, so a structural equation modeling approach was used to reanalyze the data to see if a better interpretation could be found. The model derived from the Buck/Ross study was used as the baseline model. On the first run, a significant chi-square (.0001) and a low Comparative Fit Index (.81) led to a rejection of that model. A second model was tried, this one based upon the conclusions of the Buck (1992) study: two correlated factors each being measured by four variables. The initial run produced a significant chi-square (.000001) but a higher Comparative Fit Index (.91). This was deemed as good enough to warrant further investigation and the Lagrange multiplier test statistics were analyzed.

The results of the Lagrange mulitplier test indicated that there was a significant undrawn path between Listening: Multiple-choice (ML) and the Reading Factor (R). There was also a significant undrawn path between Listening: Gap-filling (GL) and the Reading Factor (R). The last significant undrawn path indicated was between Reading: Short Answer Comprehension (SR) and the Listening Factor (L). None of these indicated paths fit the theory and they were all rejected. The Lagrange multiplier test was then examined to see if there were any correlations between the error terms that needed to be added. Results indicated that there were seven significant undrawn correlations between the error terms: E5,E1; E8,E1; E6,E1; E6,E3; E8,E2; E5,E2; and E5,E3.

The interactions proposed here were too complex for the hypothesized model, suggesting that further work was needed with the measurement model and leading to a rejection of the two factor model as well.

In conclusion, the reanalysis of Buck's (1992) study offered very little evidence, either for or against the causal model. However, using the causal model, it is possible to explain why things may have gone wrong with this study. Buck pointed out that there were several problems with the measurement instruments. Besides low reliability with ML and MR, many of the students did not have enough time to complete all of the items on the reading tests. This suggests that RL was playing a strong role. If the presentation times for the texts varied, and if there wasn't enough time for the subjects to finish, then RL was probably a confounding factor which made it impossible to find either a one factor or a two factor solution. This can be avoided in future research if rate of presentation is controlled so that it is equal both within and between the measures. This could have the effect, however, of reducing the chances of finding any significant difference between listening and reading comprehension.

Griffiths (1992) presented a study on the relationship between speech rate and listening comprehension which is relevant to Carver's causal model and rauding theory because Carver (1977, 1984, 1985a, 1985b, 1990b) has proposed mathematical relationships between an individual's rauding rate (RR), the rate of passage presentation (R), the rauding accuracy level of the individual (AR), and the accuracy of passage comprehension (A). If three of these

variables are known about an individual, then the fourth can be estimated. In Griffiths' study, speech rate (R) was manipulated to see what effects it had upon listening comprehension (A).

Griffiths had 24 young adult Japanese Elementary School teachers taking part in a university in-service training program listen to three different texts (difficulty level = 3.5 as measured in GE units) at three different speeds (fast = 250 wpm, average = 189 wpm, and slow = 127 wpm). Text and speed were controlled in a randomized complete-block design to minimize text and speed interactions. Accuracy of comprehension was measured with a 15 question true-false test given immediately after listening to the passage. K-R20 was used to estimate the reliability of the tests, although it was not clear if the .80 reliability coefficient reported was for one test or the average of all the tests. In addition, the validity of the reliability coefficient is questionable due to the fact that it seems to have been obtained from the responses of the native speakers who validated the test, not from the actual scores obtained from the subjects.

According to rauding theory there should be a decrease in accuracy of comprehension (A) as rate of presentation (R) increases. This prediction was confirmed by the results of Griffiths' study. At the slow listening speed, students comprehended 80% of the material. At the average listening speed they comprehended 68% of the material. At the fast listening speed they comprehended 62% of the material. That is, there was a constant decline in comprehension as speed was increased.

Accuracy of comprehension (A) can be expressed mathematically in rauding theory if an individual's Rauding Accuracy (AR), Rauding Rate (RR) and Rate of Presentation (R) are known (Carver, 1977, 1997). Equation (6) below is hypothesized to be valid when the rate of presentation is faster than or equal to the rauding rate of the individual (R > RR).

$$A = AR\ (RR)\ (1 / R) \tag{6}$$

Where

A = percentage correctly answered on a test of comprehension

$$AR = .04\ (AL - DL) + .64 \tag{7}$$

RR = the rauding rate of the individual, measured in Spm (standard length sentences, 1 SPM equals 16.67 standard length words, or 100 character spaces)

R = presentation time, or the amount of time the individual is permitted to read the text, measured in Spm

Equation (6) states that the amount an individual comprehends of a given passage depends on three factors: (a) the difference between the Accuracy Level (AL) of the individual and the difficulty of the text (DL); (b) the rate at which the individual reads the passage (RR); and (c) passage presentation time, or the

amount of time allowed for the individual to read (R). Equation (7) states that if the accuracy level of an individual, AL, (measured in GE units) is the same GE unit as the text, DL, then the individual will understand 68% of the complete thoughts in the text when they read it once at their normal reading speed. That is, if an individual whose AL is at the 12th grade level reads a text at the 8th grade level, that person will understand 84% of the complete thoughts in the text when it is read once at the individual's normal reading rate (AR = .04(12 - 8) + .64 = .84).

This relationship stated in equation (7) between rauding accuracy level (AL) and the relative easiness of the text (AL - DL), was determined empirically by Carver (1990b). Notice that AR depends entirely on an individual's accuracy level (AL) and the relative difficulty of the text (DL). Notice also that when the rate of presentation is controlled so that RR = R, A depends on AR and AR depends on AL and DL. If R is manipulated so that R > RR (as occurs in Griffiths' study), there will not be enough time for individuals to read the passage once at their rauding rate (RR), and equation (6) predicts a loss in comprehension that increases linearly as the difference between R and RR increases. In other words, if the same individuals above had a RR of 20 Spm (333 Wpm), and the text was presented at a rate of 25 Spm (475 Wpm), then they would only understand 67.2% of the complete thoughts in the passage (A = .84(20)(1 / 25) = .672). Note that relationship expressed in equation (6) is linear because rauding theory holds

that people's rauding rates (RR) are constant when they are doing normal reading.

Griffiths' study did not report either the ability level (AL) of the students or their rauding rates (RR). Assuming an AL and RL of 6.5, we can use the equations from rauding theory to reanalyze Griffith's data and predict the students' accuracy of comprehension (A). Note that different values could be substituted here without changing the results. First, AR must be solved.

$$AR = .04\,(6.5 - 3.5) + 0.64 = .76 \qquad (8)$$

Equation (8) shows that the group of students would have understood 76% of the text had they read it once at their rauding rate. Next, the rates of presentation (R) need to be determined. Since all of the texts are at the same difficulty level, we can assume that average word length is the same for all three texts; therefore there should not be a problem converting the wpm units into standard length words (Wpm). Remember that one standard length word equals 6 character spaces and one standard length sentence equals 16.67 Wpm. Thus, R for the fast presentation speed is 127 Wpm or 7.66 Spm; R for the average presentation speed is 188 Wpm or 11.26 Spm, and R for the fast presentation speed is 250 Wpm or 15 Spm.

As stated earlier, the students' hypothesized average rauding rate (RR) is 6.5 GE, which translates into 175 Wpm or 10.5 Spm. Now, equation (6) can be

used to predict the accuracy of comprehension for both the fast presentation speed (250 wpm) and the average presentation speed since in both cases R is greater than RR (15 > 10.5 and 11.26 > 10.5). For the fast presentation speed A is predicted to be

$$A = (0.76)\ (10.5) / (1 / 15) = .53 \tag{9}$$

The actual A reported by Griffiths was .62, which is a 9% difference. For the average presentation speed, A is predicted to be

$$A = (0.76)\ (10.5) / (1 / 11.26) = .71 \tag{10}$$

The actual A reported by Griffiths was .68, which is a 3% difference.

For the slow presentation- speed (127 wpm or 7.66 Spm), equation (6) is not valid because R < RR (7.66 < 10.5). When R < RR, A is calculated as follows:

$$A = (1 / R) / [(1 / R) + i] \tag{11}$$

where

$$i = (1 / RR)\ [(1 / AR) - 1] \tag{12}$$

The “i” in equations (11) and (12) represents the inefficiency constant; it is the amount of inefficiency associated with having more than enough time to read a passage once at the rauding rate (Carver, 1977). If the individual from the example stated earlier read the same text, but this time the rate of passage presentation was 10 Spm (167 Wpm), the value of i can be calculated as: i = (1 / 20)[(1 / .84) - 1] = 0.01. Then A can be calculated as: A = (1 / 10) / [1 / 10) + 0.01] = 0.91. Therefore, if the individual read the passage twice at their normal reading rate, equation (11) predicts that this person would comprehend 91% of the complete thoughts in the text. Note that when R<RR, the relationship between A and R is curvilinear.

For the slow presentation speed in Griffiths' (1992) study, R is 7.66 and therefore less than the hypothesized RR of 10.5 stated earlier. Therefore, equation (11) must be used to predict A. First, in equation (12) i is calculated as,

$$i = (1 / 10.5)[(1 / .76) - 1] = .03007 \qquad (13)$$

Then, equation (6) can be used to predict A for the slow presentation-speed condition (7.66 Spm).

$$A = (1 / 7.66) / [(1 / 7.66) + .03007] = .81 \qquad (14)$$

The reported A for the slow presentation-speed was .80, which is only 1% different from the predicted value. Considering that the reported reliability coefficient for the comprehension tests was .80, it seems fair to assume that the 9% difference between the actual and predicted A values in the fast condition, the 3% difference in the average condition, and the 1% difference in the slow condition are well within the standard error of measure for that test.

In conclusion, equations (6), (7), (11) and (12) predicted A with a remarkable degree of accuracy. It should be noted, however, that there was not one unique solution for any of these equations. The ability level (AL) and the rauding rate (RR) of the subjects were both assumed to be equal (6.5 GE units). Similar estimates of A could have been obtained by using a higher estimate of AL with a lower estimate of RR. Since neither AL nor RR was reported and they both had to be assumed, this reanalysis of Griffiths' data only provides limited support for rauding theory and the validity of equations (6), (7), (11) and (12). Were this study to be replicated, it would be vital to measure students' AL and RL so that A could be predicted. In addition, a larger *n*-size would add great strength to any findings.

In 1997, Zhao investigated the effect of speech rate (R) on listening comprehension (A). While Zaho's study suffers from numerous problems, including low *n*-size and lack of any reported reliability coefficients for the comprehension measures, it will be discussed here because it is related to the relationships postulated by equations (6), (7), (11) and (12). Subjects were 15

young adults, 10 of which were enrolled in an intensive English language program at a major US university, and five of which were graduate students at the same university. Ability levels ranged from intermediate to advanced.

Zhao had subjects listen to four passages taken from the listening section of the TOEFL. Each passage was subject to a different condition. In Condition 1, subjects could not control either speech rate or repetition. In Condition 2, subjects could control rate but not repetition. In Condition 3, subjects could control both rate and repetition. Finally, in Condition 4, students could control repetition but not rate. Results indicated that subjects had the highest comprehension when they could control rate, and the difference between the rate-controllable conditions (2 and 3) and rate-non-controllable conditions (1 and 4) was significant at $p < .01$.

When subjects could control rate, they invariably chose a rate slower than the original presentation rate of 192 wpm. The mean chosen rate speed was 128 wpm, leading Zhao to conclude “that slower speeds improved listening comprehension and are thus consistent with the conventional wisdom concerning the relationship between speech rate and comprehension: the slower the speed, the better the comprehension” (Zhao, 1997, p. 60).

In conclusion, Zhao’s study would have been strengthened greatly by the addition of measures of AL, DL, RR. The *n*-size (15) was too small for parametric statistics, and without reporting reliability coefficients for the measures of comprehension it is impossible to know with certainty if the differences found

were statistically significant. Nevertheless, the study's final results, that comprehension increases as the rate of presentation is slowed, provides limited support for equations (6), (7), (11) and (12) above.

In 1998, Bae and Bachman investigated the factorial distinctness of reading and listening with students in grades two, three, and four. This study is directly relevant to Carver's causal model because it investigates differences in efficiency of listening and reading with beginning readers. All of the children were enrolled in a Korean/English two-way immersion program, and they were separated into two groups: (a) children of Korean-American ancestry (*n* = 120), and (b) non Korean-American ancestry (*n* = 36).

For advanced readers the causal model predicts no difference between listening efficiency and reading efficiency because PL is not considered to affect AL for advanced readers. For beginning native speaker readers, however, PL is considered to have a causal effect on AL. According to the causal model, Listening efficiency (EL) equals the average of an individual's verbal level (VL) and rate level (RL). Reading efficiency (EL), on the other hand, equals the average of accuracy level (AL) and RL. Since PL is hypothesized to have a causal effect on AL but not VL, it is logical that for beginning native speaker readers, listening comprehension and reading comprehension would appear as separate factors because PL is affecting AL but not VL. For beginning non-native speaker readers, the causal model does not make any predictions about differences in listening and reading efficiency level.

Listening comprehension was measured with three different tasks. Listening Task 1 (Ltask1) attempted to measure students' ability to comprehend spoken Korean involving concepts associated with daily life and elementary school subjects. Reliability in the form of a squared multiple correlation was .61. Listening Task 2 (Ltask2) tried to measure students' ability to comprehend spoken instructions asking them to draw lines or mark spaces. Reliability in the form of a squared multiple correlation was .36. Listening Task 3 (Ltask3) asked students to listen to longer passages and then answer three one sentence comprehension questions. Reliability in the form of a squared multiple correlation was .73.

Reading comprehension was also measured with three different tasks. Reading Task 1 (Rtask1) asked students to circle words or objects that corresponded to the written language prompts next to them. Reliability in the form of a squared multiple correlation was .68. Reading Task 2 (Rtask2) had students read longer, more complex prompts and then give short answers in writing, either in English or Korean. Reliability in the form of a squared multiple correlation was .61. Reading Task 3 (Rtask3) asked students to read eight longer passages, each followed by four comprehension questions. Reliability in the form of a squared multiple correlation was .74.

A correlated two factor model was hypothesized to best represent the data. Separate analyses were performed on each group. For the Korean American group (n = 120), a non-significant chi-square (.489) and a Comparative

Fit Index of 1.0 indicated a good fit. The two factors correlated at .86. For the non-Korean American group (n = 36) a non-significant chi-square (.165) and a Comparative Fit Index of .93 also indicated a good fit. For the non-Korean American group, the two factors correlated at .77. Because there was a high correlation between the two factors, an alternative model with a single factor was then tested to see which fit the data better. The single factor model had a chi-square of .001, which indicated that the correlated two factor model was better.

Bae and Bachman (1998) concluded that listening and reading were "separable" factors, but noted that the high correlation between the factors suggested that they were very similar processes. This conclusion is highly consistent with the causal model's predictions for native speaker performance; AL did not equal VL because of the influence of PL. However, it is unclear why the same results were found for the non-native speaker group. For this group AL probably did not equal VL, but it seems illogical to assume that this was because of the influence of PL. An alternative explanation is that RL became a factor with the non-Korean American group's performance on the listening tests. Unfortunately, no measure of RL or PL was used in the study, making the results inconclusive.

In the end, however, we have to ask if it is worthwhile to consider two factors that correlate at .80 to be different. Most researchers would consider a correlation of .80 between parallel forms of a test to be high enough to assume both forms were measuring approximately the same factor. Nevertheless,

regardless of whether we accept the one factor or the two factor solution, these results lend moderate support for the validity of the causal model being applied to ESL situations.

Hirai (1999) investigated Carver's (1982) claim that an individual's optimal reading rate is the same as their optimal listening rate. Unlike Carver, however, Hirai's subjects were 56 Japanese college students studying English as a foreign language, making this study extremely relevant to an investigation of the applicability of the Rauding model to non-native speakers of English.

Hirai's research started with three basic hypotheses. First, there would be a strong correlation between proficiency and the optimal reading rate. Second, there would be a strong correlation between proficiency and the optimal listening rate. Finally, the optimal listening and reading rate would be the same.

Hirai's definition of optimal rate was based on the one given in Carver's (1982) research, the highest speed at which 75% accuracy could be obtained. This is, by definition, their rauding rate (see law three of rauding theory above). Unfortunately, more than half of the participants never achieved 75% accuracy, so the working definition of optimal rate was changed to "most efficient rate," which was calculated by equation (1), E = AR.

These hypotheses are grounded in the three laws of rauding theory. Hypothesis one and two originally stated that RL would correlate with EL, which is a fundamental assertion of rauding theory and the rauding model. The modified versions, however, simply stated that a measure of efficiency (E) would

correlate with their efficiency level (EL). In other words, (R)(A) = (RL + AL) / 2. That means that a measure of an individual's performance (R)(A) should equal a measure of that individual's ability (RL + AL) / 2. Hypothesis three stated that EL (or E) is the same, no matter if the individual is listening or reading. Put differently, a measure of an individual's best performance, defined as (A)(R), should be the same in either reading or listening. These three hypotheses are therefore based on the relationships outlined in equations (1), (2), (3), (4) and (5).

Proficiency was measured with a 50-item cloze test taken from a passage in an English textbook in which every seventh word had been deleted. Exact scoring was used and one point was given for each correct answer. Cronbach's alpha for the test was .90.

Reading efficiency was measured with two short texts, each having eight multiple-choice questions. In each of the two trials, students' reading times were measured and then they took the eight-question comprehension test. The trial that had the best efficiency measure (calculated by equation (1), E = AR) was considered their optimal reading efficiency. Cronbach's alpha was .70 for the tests, but the appropriateness of this statistic is dubious since it was calculated by merging both tests together to form one longer sixteen item test, even though in the end only one test was used to calculate the efficiency measure.

Listening efficiency was measured with five short passages, each having eight multiple-choice questions. The rate of passage delivery was controlled with

a speech compressor and it increased approximately 30 Wpm per trial. The participants' optimal listening rates were calculated by selecting the trial that had the highest efficiency rating, calculated by equation (1), E = AR. Cronbach's alpha was .89 for the listening comprehension tests, but the appropriateness of this statistic is once again dubious since it was calculated by merging all five tests together to form one longer forty item test, even though in the end only one test was used to calculate the efficiency measure.

Correlation coefficients were calculated for the cloze test and the measures of optimal listening and reading efficiency. There was a .74 correlation between the cloze test and listening efficiency. There was a .72 correlation between the cloze test and reading efficiency. Finally, there was a .94 correlation between listening and reading efficiency.

These results indicate moderate support for equations (1), (2), (3), (4) and (5) as well as indirect support for the three laws of rauding theory. This is especially true when one considers that the already high correlation between listening and reading efficiency would probably have been even higher had the tests been more reliable. Indeed, with a larger *n* size and more reliable instruments, this experiment could have strongly supported the assertions of rauding theory. Nevertheless, the results suggest that efficiency of comprehension, whether it be in listening or reading, may be the same for non-native speakers of English. This would imply that any time a comparison was

made between comprehension in listening and reading, it would be vital to account for differences in presentation time.

In 1999, Nassaji and Geva administered 15 reading related measures to 60 Iranian ESL graduate students. These were measures of reading comprehension, word recognition, reading rate, phonological processing skill, orthographic processing skill, syntactic processing skill, semantic processing skill, working memory and cognitive speed. Since virtually all of these variables are part of rauding theory and the causal model, this study is directly related to the validity of the causal model with non-native speakers of English.

Nassaji and Geva analyzed their data with correlational and hierarchical multiple regression. For the regression analysis, reading comprehension, silent reading rate, and word recognition were used as the dependent variables. All other variables were used as independent variables. Of these, phonological processing skill, orthographic processing skill, syntactic processing skill, and semantic processing skill were separated into three different variables: one reflecting the accuracy of the process, another the speed, and then the their *z* scores were averaged to form an efficiency measure.

Nassaji and Geva summarized the results of their correlational analysis by stating that “there was a positive and significant relationship between lower level cognitive and linguistic variables and adult ESL reading” (1999, p. 254). This was based upon the high correlations between the efficiency measures for phonological processing skill, orthographic processing skill, and reading

comprehension. This is consistent with the causal model because the efficiency measures were comprised of measures of PL (which is a component of AL) and RL, both of which should correlate highly with EL (reading comprehension).

Another finding was that cognitive speed did not correlate with any of the dependent variables. Nassaji and Geva said, “these data seem to suggest that speed of letter naming as measured by RAN, while not wholly related to adult L2 reading comprehension, may be related to subcomponents of L2 reading comprehension” (1999, p. 254). The causal model, however, predicts that CS should correlate strongly with reading rate because it is a measure of RL. This finding did not match the causal model. However, they reported that CS did correlate with the efficiency scores of the phonological, orthographic, syntactic, and semantic processing measures. The causal model considers efficiency measures such as these to be measures of EL and therefore comprised of RL and AL (Carver, 1999). Carver (1992d) presented factor analysis data of similar measures showing that efficiency variables can load almost as highly on a rate factor as they do an efficiency factor. The measures of rate in that study had much more variance than the measures of accuracy. Carver argued that this finding explained why the EL measures created by combining accuracy scores and rate scores loaded higher on a rate factor. Since CS is hypothesized to be the major cause of RL with adult readers, the finding that CS correlated highly with measures of efficiency in Nassaji and Geva’s study can be considered consistent with the causal model.

Finally, Nassaji and Geva reported that “both speed and accuracy on L2 phonological, orthographic, syntactic, and semantic measures correlate significantly with each other and with various indices of L2 reading proficiency” (1999, p. 260). This finding is also consistent with the causal model, since RL and AL are considered to be correlated (Carver, 1992d, 1999; Carver & Darby, 1972).

In their multiple regression analysis, Nassaji and Geva reported that seven variables (cognitive speed, working memory, phonological efficiency, orthographic efficiency, syntactic efficiency, semantic efficiency, and vocabulary) could explain 56% of the variance in reading comprehension. The same variables explained 47% of the variance in silent reading rate and 45% of the variance in word recognition. They concluded from this that “efficient lower level phonological and orthographic processing skills as well as higher level syntactic and semantic skills contribute significantly to various indices of ESL reading” (1999, p. 260). Again this is consistent with the causal model because PL (phonological and orthographic processing skills) directly effects RL and AL, which are the fundamental components of EL and therefore highly related to syntactic and semantic processing skills.

The analysis and subsequent conclusions made by Nassaji and Geva are congruent with the causal model, providing moderate support for the validity of the causal model with non-native speakers of English. The only finding that was not predicted by the causal model was the lack of any correlation between silent

reading rate and cognitive speed. This suggests that this aspect of the causal model may need to be modified for non-native speakers of English.

Despite the fact that the measures Nassaji and Geva used are the important constructs in the causal model, they did not analyze their data from the perspective of rauding theory or the causal model. Therefore their data will now be reanalyzed from the perspective of the causal model. This reanalysis will include factor analysis and structural equation modeling using a covariance matrix constructed from the correlation matrix included in their study. The variables, their reliability coefficients and their hypothesized constructs in the causal model are as follows.

Reading comprehension (R-Comp) was measured with the Reading section of the Nelson-Denny reading test and had a reliability coefficient of .85. It is hypothesized to be an measure of EL.

Silent reading rate (Rate) was determined on the basis of the first minute reading of the first passage of the Nelson-Denny reading test. It is hypothesized to be a measure of RL. No reliability measure was reported for Rate.

Single word recognition (Word) was measured by the word reading section of the Wide Range Achievement test (WRAT-3). It is hypothesized to be a measure of PL and had a reliability coefficient of .86.

Phonological processing skill (Pho) was measured with a pseudo-word-matching task and is therefore a measure of decoding knowledge (DK). Subjects looked at two pseudo-words (e.g., flumb-flum) and decided if they had the same

sound or not. In the causal model, the construct PL can be expressed mathematically as (DK + AL)/2; therefore Pho is a measure of DK and a fundamental component of PL. The accuracy of subjects' responses as well as how long it took them to answer all the items were recorded. The time to respond to all the items was made into a variable called Pho-t. Pho-t is hypothesized to be a measure of RL. No reliability coefficient was reported for Pho-t. In addition, an efficiency measure (Pho-eff) was also created by combining Pho and Pho-t. This variable is hypothesized to be a measure of EL. No reliability coefficient was reported for Pho-eff.

Orthographic processing skill (Orth) was measured with a task that required subjects to look at pairs of non-words (e.g., filve-filv; gmub-gnub) and decide if either of the two words did not conform to the orthographic rules of English. Orthographic knowledge is hypothesized to be a correlate of either PL or AL (Carver, 1999). Reliability for the accuracy of this task was reported to be .70. The time to respond to all the items was made into a variable called Orth-t. Orth-t is hypothesized to be a measure of RL. No reliability coefficient was reported for Orth-t. In addition, an efficiency measure (Orth-eff) was also created by combining Orth and Orth-t. This variable is hypothesized to be a measure of EL. No reliability coefficient was reported for Orth-eff.

Syntactic processing skill (Syn) was tested by asking subjects to judge the grammaticality of 30 sentences. Word frequency was controlled so that vocabulary knowledge would not adversely effect the results. Syn does not

correspond to any construct in the causal model, but was hypothesized to be a correlate of AL because it is a function of comprehension and there was no time limit on the test. Reliability was reported as .68. The time to respond to all the items was made into a variable called Syn-t. Syn-t is hypothesized to be a measure of RL. No reliability coefficient was reported for Syn-t. In addition, an efficiency measure (Syn-eff) was also created by combining Syn and Syn-t. This variable is hypothesized to be a measure of EL. No reliability coefficient was reported for Syn-eff.

Semantic processing skill was measured with two tasks. One was a test of lexical knowledge (Lex-sem). The vocabulary section of the Nelson-Denny reading test (form F) was used to measure Lex-sem, which has been used in prior research to measure AL (Carver, 1992c). Reliability was reported as .84. The other measure of semantic processing skill was a test requiring subjects to judge the semantic goodness of sentences such as “A timid accident devastated a huge crop” and “The angry teacher punished the rude student.” This is a test of comprehension and it did not have a time limit; therefore it is hypothesized to be a measure of AL. Reliability was reported to be .67. The time to respond to all the items was made into a variable called Sem-t. Sem-t is hypothesized to be a measure of RL. No reliability coefficient was reported for Sem-t. In addition, an efficiency measure (Sem-eff) was also created by combining Sem and Sem-t. This variable is hypothesized to be a measure of EL. No reliability coefficient was reported for Sem-eff.

Cognitive speed was measured with a test called the Rapid Automatization Naming Test (RAN), which requires subjects to read a randomized string of letters as fast as they can. This is the same test used in rauding theory to test cognitive speed (CS). Reliability was not reported.

According to Carver (1977, 1992a, 1992b, 1999), measures of AL, RL, and EL will always form one factor (EL) or two factors (AL and RL). Following the procedure outlined in Carver (1992c, 1992d), two principal components analyses were used on the variables hypothesized to be measures of EL, AL, and RL from Nassaji and Geva's (1999) study. In Analysis 1, the number of eigen values greater than 1.0 was used to determine the number of factors. Since Analysis 1 resulted in a two-factor fit, Analysis 2 forced a single-factor fit. When there were two factors, an oblique rotation (direct oblimin) was used because prior research (Carver, 1992b; Carver, 1992d; Carver & Darby, 1972) showed that AL and RL were correlated. The best measure of each factor was empirically determined by the highest loading on that factor and that measure was used to name the factor.

Table 2 contains the results of the two factor analyses described above. In Analysis 1 two factors resulted; they correlated .44. It is obvious that the first factor was a Rate factor because it loaded highly on the variables that purportedly reflect rauding rate level (RL) with the notable exception of Rate. The highest loading was on Pho-t, which was hypothesized to measure RL. Surprisingly, Pho-eff had a .69 loading, which was higher than .67 for Syn-t. This

was unexpected, but is probably related to the fact that Pho-t had the highest loading on this factor.

The second factor was obviously an Accuracy factor because it loaded highly on the variables that purportedly measure rauding accuracy level (AL). The highest loading was on S-sem, which was hypothesized to measure AL. However, the relatively low loading on Lex-sem, .56, was a surprise. As a result, both R-Comp and S-sem-eff loaded higher on the Accuracy factor than did Lex-sem despite that they were hypothesized to be measures of EL, EL, and AL respectively. In addition, Rate (.56) loaded about the same as Lex-sem on the Accuracy factor.

Analysis 2 in Table 2 is a single factor fit forced upon the data. This factor was obviously an efficiency factor because it loaded highly on the variables that purportedly measured rauding efficiency. The highest loading was on Syn-eff, which was hypothesized to reflect EL. The relatively low loadings of Pho-eff (.68) and Orth-eff (.66) were a surprise. In addition, the high loading of Lex-sem (.75) on the Efficiency factor was a surprise.

In Table 2 it should be noticed that one of the three loadings for each of the thirteen variables has been shaded. This shading marks the loading that was predicted from rauding theory to be the highest of the three loadings for that variable. Note that this prediction was correct in 10 of the 13 cases. The exceptions were: (a) Rate (RL), which hardly loaded on RL (-.21) and had its highest loading on the Efficiency factor (.69); (b) Lex-sem (AL) which loaded

higher on the Efficiency factor (-.75) than it did on the Accuracy factor (.56); and (c) Pho-eff (EL), which loaded higher on the Rate factor (.69) than it did on the Efficiency factor (.68).

The first exception suggests that there may be some problems with the construct of rauding Rate level and silent reading rate. This problem was hinted at in Nassaji and Geva's original analysis of the data where they pointed out that cognitive speed (CS) did not significantly correlate with Rate. The second exception suggests that other factors besides vocabulary knowledge contribute to rauding accuracy level. This may be because vocabulary knowledge was related to silent reading rate, which would explain: (a) why they both loaded so highly on the Efficiency factor; and (b) why Rate loaded so highly on the Accuracy factor. The measure of Rate used in this study was taken by counting the number of words read in one minute on the Nelson-Denny reading test. If this

Table 2. Results of the Factor Analysis of 13 Variables That Purportedly Measure Individual Differences in Rauding Accuracy Level, Rauding Rate Level and Rauding Efficiency Level

		Analysis 1		Analysis 2
Variable	Theoretical Correlate	FACTOR 1 (Rate Factor)	FACTOR 2 (Accuracy Factor)	FACTOR 3 (Efficiency Factor)
S-sem-t	RL	0.7833	-0.0089	.6923
Rate	RL	-0.2199	0.5688	-.6484
Pho-t	RL	0.8382*	0.0903	.6607
Orth-t	RL	0.8274	0.0880	.6531
Syn-t	RL	0.6725	-0.0880	.6588
S-sem	AL	0.3085	0.9210*	-.4688
Lex-sem	AL	-0.3468	0.5671	-.7582
Syn	AL	0.0207	0.7295	-.5669

R-Comp	EL	-0.2313	0.6305	-.7079
Pho-eff	EL	0.6953	-0.0997	.6881
Orth-eff	EL	0.6112	-0.1582	.6615
Syn-eff	EL	0.4287	-0.5436	.8109*
S-Sem-eff	EL	0.3249	-0.6545	.8091

Note: The asterisk indicates the highest loading for that factor.

reading section had unknown words in it for many of the subjects, it would explain why vocabulary knowledge was related to rate and therefore efficiency.

The third exception can be explained simply by the fact that Pho had a standard deviation of 3.89 (*M* = 22.58, range = 15 - 30), which is relatively little variance. The amount of variance for Pho-t, on the other hand, was quite high (*SD* = 26.7, *M* = 97.42, range = 60 - 170). It is no surprise that Pho-eff was more influenced by the variance in Pho-t than in Pho, which could easily result in a higher loading on the Efficiency factor.

In conclusion, this reanalysis of Nassaji and Geva's (1999) data provides strong correlational support for the existence of one factor (EL) or two factors (RL & AL) in non-native speaker reading comprehension. However, it also raises two important questions. First, why did silent reading rate (Rate) not load on the Rate factor, instead loading much higher on the accuracy and efficiency factors? Second, why did vocabulary knowledge (Lex-sem) load higher on the efficiency factor than it did on the Accuracy factor? These results, along with earlier data that showed that RAN (a measure of CS) did not correlate with Rate, suggest that a closer look needs to be taken at the Rate Level construct (RL) and the Accuracy Level construct (AL) with non-native speakers of English.

In the causal model, with advanced native speaker readers, Rate Level is hypothesized to be caused primarily by cognitive speed. However, Nassaji and Geva (1999) concluded that there was not any relationship between silent reading speed (Rate) and cognitive speed (RAN) with their 60 advanced ESL readers. However, they did find a relationship between cognitive speed and the efficiency of phonological, orthographical, syntactic, and semantic processes. Along with reading rate, these processes are all purported to be measures of rate level; according to rauding theory they should all load together on one factor. Clearly this needs to be investigated further.

The prior factor analysis in Table 2 supports this conclusion as well; reading rate only had a -.22 loading on the rate factor. Instead, it loaded .57 on the Accuracy and -.64 on the Efficiency factors. Another anomaly was that Lex-sem loaded higher on the Efficiency factor than it did on the Accuracy factor. Could this be related to Rate's unusual behavior? One possible interpretation of the results is that Lex-sem was affecting Rate. This would account for Rate's high loading on Accuracy as well as Lex-sem's higher loading on efficiency. It would also explain why the measure of cognitive speed (RAN) did not correlate with Rate. The number of unknown words on the first passage of the Nelson-Denny Reading test could easily have a strong impact on Rate. The more unknown words there were, the slower individuals would be expected to read and the less likely they would be to read at their rauding rate. Instead, they would be much more likely to slow down and use a studying process, which is slower

and is not hypothesized to be controlled by CS like the rauding rate is. Furthermore, the reanalysis of Oller's (1973) study above found a strong relationship between silent reading rate and vocabulary, which provides further support for the hypothesis that silent reading rate may load on an accuracy factor rather than a rate factor.

A third factor analysis was performed to investigate these questions. Variables included were RAN (CS), Pho-t (RL), Orth-t (RL), Syn-t (RL), Rate (RL), Sem-t (RL), Lex-sem (AL), and S-sem (AL). According to the causal model, this should result in two correlated factors, a speed factor and an accuracy factor, with RAN loading on the speed factor because CS is a proximal cause of RL. However, it was hypothesized that for the 60 ESL subjects, that the results would be slightly different. If vocabulary significantly affected silent reading rate, then Rate would load higher on the Accuracy factor than the rate factor.

Table 3 shows the results of the factor analysis. Two factors were found with eigen values greater than 1.0 and an oblique rotation (direct oblimin) was used because the prior factor analysis showed that AL and RL were correlated.

Table 3. Results of the Factor Analysis of Eight Variables That Purportedly Measure Individual Differences in Rauding Accuracy Level and Rauding Rate Level

Variable	Theoretical Correlate	FACTOR 1 (Speed Factor)	FACTOR 2 (Accuracy Factor)
Syn-t	RL	0.6075	-0.2223
RAN	CS	0.5951	0.1231
Pho-t	RL	0.7581	0.0045
Orth-t	RL	0.7785	-0.0025

S-sem-t	RL	0.7838	-0.0662
S-sem	AL	0.2043	0.7683
Lex-sem	AL	-0.3520	0.5882
Rate	AL	-0.2313	0.6260

In this analysis they correlated at .28. Factor 1 is clearly a speed factor as the highest loadings are all with variables that are purportedly measuring RL. Factor 2 is clearly an accuracy factor as the highest loadings are all with variables purported to measure AL. Note that one of the two loadings for each of the eight variables has been shaded. This shading marks the loading that was predicted to be the highest of the three loadings for that variable. Note that this prediction was correct in all cases.

In conclusion, the fourth factor analysis provided strong correlational evidence that CS is a component of RL even for non-native speakers of English. In addition, it provided strong evidence that measures of silent reading rate may not reflect RL even with advanced level ESL readers. Results indicated that measures of silent reading rate were more related to AL than RL. These results, however, do not invalidate applying the causal model to non-native speaker data. Encountering unfamiliar words probably forces readers to use a process other than the rauding process when they read, for instance a studying process. If an individual is studying instead of rauding, then there would be no reason for CS to correlate with their reading rate, nor for their reading rate to correlate with other measures of RL. Therefore, these results are congruent with the causal model.

Accuracy Level (AL) in rauding theory is hypothesized to have two proximal causes for beginning and intermediate readers: Verbal Level (VL) and Pronunciation Level (PL). It can be represented mathematically as AL = (VL + PL) / 2. This means that the number of words that an individual can read and understand (AL) depends on (a) how many words they understand in listening (VL) and (b) how well they can pronounce words that are written (PL). For native speakers, who posses a large oral language base before they begin to read, this is a logical argument. However, many of these assumptions are invalid for non-native speakers. Most non-native speakers, for instance, do not possess a large oral language base before they begin to read.

Pronunciation level (PL) in rauding theory is hypothesized to have two proximal causes with beginning and intermediate readers: decoding knowledge (DK) and accuracy level (AL). It can be represented mathematically as PL = (AL + DK) / 2. This means that the number of words individuals can pronounce in a list (PL) depends on how many words they know (AL) and their decoding knowledge (DK). In the causal model, PL is hypothesized to be a proximal cause of RL and AL. As stated earlier, however, it seems illogical to assume that PL could be causal with AL with non-native speakers of English. Learners of English do not start with a large oral language base, and therefore there is no reason for PL to be causal in reference to AL. Indeed, it seems logical to consider PL a highly correlated component of AL. Second language vocabulary specialists such as Nation (1990) argue that knowing the pronunciation of a word is simply

part of what it means to know the word. This assertion supports the idea that for advanced ESL learners PL and AL are the same factor. However, there is a problem with incorporating the PL construct into the AL construct. It is illogical to assume that AL and decoding knowledge (DK) are the same factor. Instead, it seems more logical to assume that AL and DK are different, but correlated factors.

Can measures purporting to measure PL, AL, and DK from Nassaji and Geva's (1999) study be broken down into only two factors, AL and DK? Since AL and DK should be highly correlated, a two analysis approach was used. In Analysis 1, the number of eigen values greater than 1.0 was used to determine the number of factors. Since Analysis 1 resulted in a single-factor fit, in Analysis 2 a two-factor fit was forced. An oblique rotation (direct oblimin) was used because the factors were hypothesized to be correlated. Table 4 below shows the results of Analysis 1 and Analysis 2.

Analysis 1 resulted in a one factor fit and the factor was clearly an Accuracy factor because the highest loading was on Lex-sem (.79) which is a measure of AL. The second highest loading was on Word (.75), followed by S-sem (.74). Analysis 1 provided moderate support for the hypothesis that AL and PL are only one factor with advanced ESL learners.

In Analysis 2, a two factor fit was forced on the data; the factors correlated at -.66. Factor 1 was clearly an Accuracy Factor because the highest loading was on Lex-sem (.74). Factor 2 was clearly a Decoding Knowledge factor,

because the only variables that loaded on it were Pho (.63) and Orth (.70). Note that in Analysis 2, one of the two loadings for each of the seven variables has been shaded. This shading marks the loading that was predicted to be the highest for that variable. Note that this prediction was correct in all cases. In

Table 4. Results of the Factor Analysis of Seven Variables That Purportedly Measure Differences in Accuracy Level and Decoding Knowledge

Variable	Theoretical Correlate	Analysis 1		Analysis 2
		FACTOR 1 (Accuracy Factor)	FACTOR 1 (Accuracy Factor)	FACTOR 2 (Decoding Knowledge Factor)
Lex-sem	AL	0.7975	0.7480	0.0011
S-sem	AL	0.7427	0.7294	-0.0396
Syn	AL	0.7194	0.6473	0.0348
Word	PL	0.7519	0.7209	-0.0184
Rate	AL	0.6966	0.6208	0.0411
Pho	DK	0.6184	0.0634	0.6392
Orth	DK	0.5696	-0.0382	0.7083

conclusion, Analysis 2 provides moderate support for the hypothesis that decoding knowledge (DK) is highly related to but separate from Accuracy level (AL) for advanced ESL learners.

Based upon the results of these two analyses, structural equation modeling was used to assess the factorial distinctness of DK and AL. Figure 5 shows the model and EQS summary statistics. Note that Syn and Rate were not included in the measurement model because (a) they had the lowest factor loadings, and (b) they were the least related to the AL construct as defined in Carver's causal model (1999). A non-significant chi-square value (.57) with 4 degrees of freedom indicated that the data fit the model and the Comparative Fit Index was 1.0, suggesting a good model fit. This model was then compared to a one factor model, but the chi-square was less than .05 and the Comparative Fit Index was .85, leading to a rejection of the one factor model.

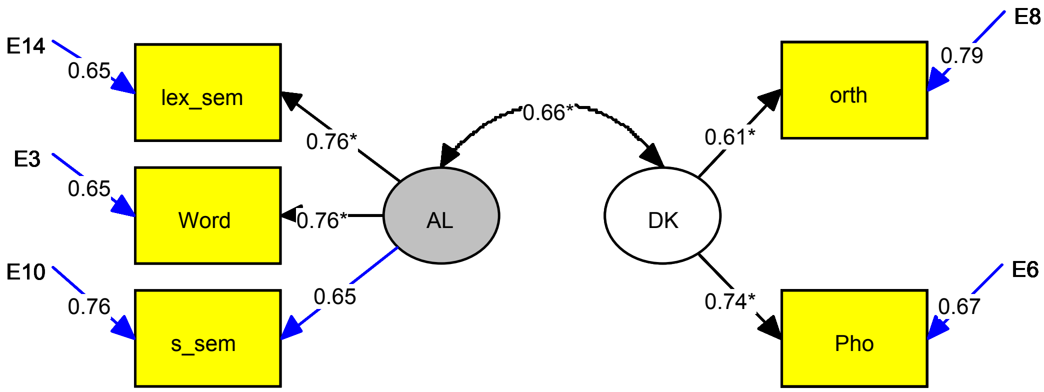

Figure 5. Structural Equation Model for Variables Purporting to Measure Accuracy Level and Decoding Knowledge from a reanalysis of Nassaji and Geva (1999) Data

In conclusion, Nassaji and Geva's (1999) study provides strong support for the applicability of the causal model to non-native speakers of English. Findings of this reanalysis can be summarized as follows: (a) variables purportedly measuring EL, AL, and RL either formed one factor (EL) or two factors (AL and RL); (b) variables purportedly measuring RL loaded on a speed factor with a measure of cognitive speed (CS), with the notable exception of Rate; and (c) variables purportedly measuring accuracy level and pronunciation level were factorably indistinguishable from one another except for measures of decoding knowledge. These conclusions provide strong correlational support for the validity of rauding theory and the causal model for non-native speakers of English. The fact that Rate loaded on AL instead of RL is not terribly detrimental to the theory because the subjects were probably reading difficult material and therefore Rate was not a measure of their rauding rate. If we assume that the

students were not rauding because there were too many unfamiliar words in the text, it would also explain why Rate loaded so highly on the Accuracy factor. It may be that in most ESL reading situations, vocabulary will always be a strong factor effecting reading rate as long as authentic native speaker level texts are read.

The fact that PL and AL were indistinguishable except for measures of DK was unexpected, but theoretically makes sense, especially if part of knowing a word for the average foreign language learner is also knowing its pronunciation. Second, most ESL learners learn the pronunciation of a word when they learn it; therefore PL should equal AL. DK not loading on AL also makes sense because they are different, but related constructs. According to the causal model, measures of PL that ask an individual to read a list of real words will only measure two things, AL and DK. This therefore supports the finding that measures of AL, PL and DK load on two factors, an accuracy factor and a decoding factor.

Based upon the above conclusions, a structural equation model was created to reflect the factor structure of the variables reported in Nassaji and Geva's (1999) study. It was hypothesized that there would be three correlated factors: AL, RL and DK. A non-significant chi-square of 0.06 and a Comparative Fit Index of .942 indicated good model fit. Figure 6 below shows the model and relevant statistics. As predicted, there was a significant correlation ($p < .05$) between RL and AL (-0.53), as there was between AL and PL (0.67). The

direction of the correlation between RL and AL was due to the fact that in this study RL was measured by counting how many seconds it took for subjects to complete their tasks. Simply put, individuals with a high AL took fewer seconds to complete the tasks, resulting in a negative correlation. The correlation between DK and RL was 0.29 and was not significant ($p > .05$). The causal model predicts a correlation between PL and RL with beginning, but not advanced readers. The participants in this study were advanced readers, but in their second language. Thus it is unclear what exactly the causal model would predict. In conclusion, these results give moderate correlational support to rauding theory and the validity of the rauding model for non-native speakers of English. Clearly the next step is to conduct more research with second language learners and the causal model.

06
25
42

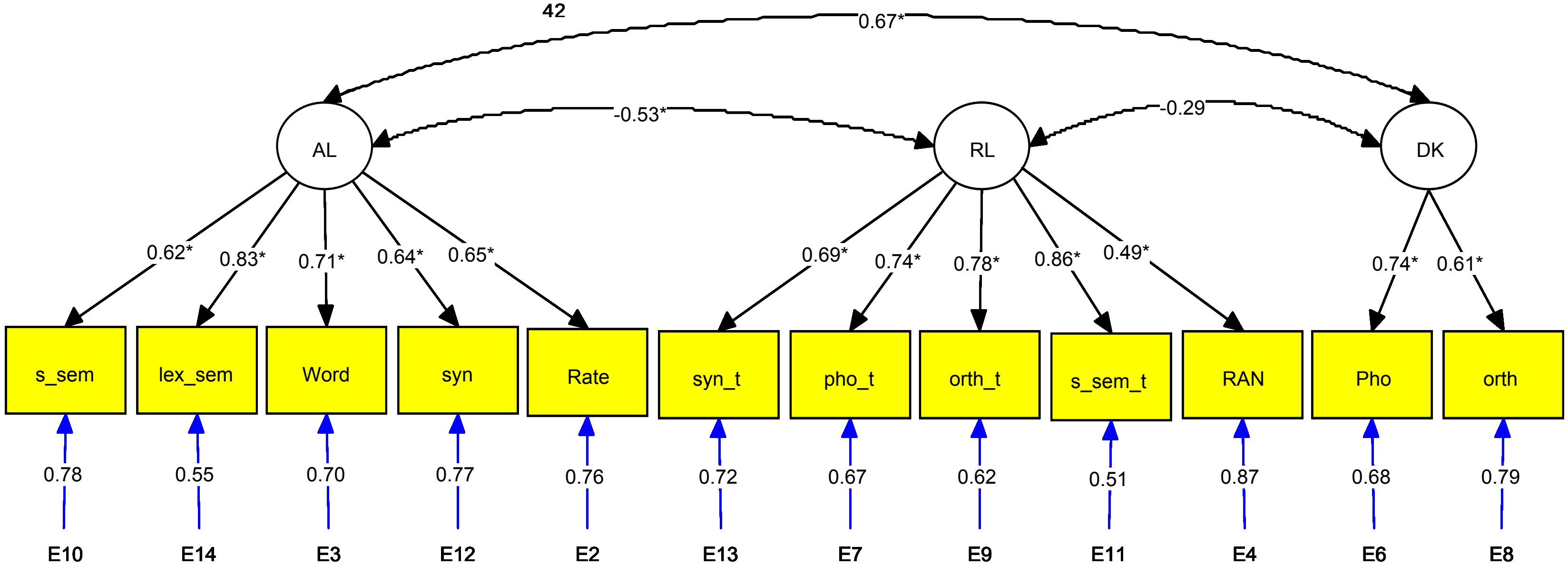

Figure 6. The Factor Structure from Nassaji and Geva's (1999) Study Measuring AL, RL and DK in Advanced ESL Reading

CHAPTER 3

METHOD

Participants

Participants were originally 130 Japanese young adults from two different universities in Tokyo. Of the original subjects, 34 were missing scores on four or more tests due to absences and were deleted from the analyses. 38% of the participants were male and 62% were female. The average age was 21.4 years old. 81 of the final subjects were enrolled in an intensive English language program (IELP) at an American University located in Tokyo. They were participating in the IELP program to prepare themselves for undergraduate study in English and their TOEFL scores ranged from 353 to 530. A total of 11 subjects came from an upper tier four year Japanese university and were participating in an intensive TOEFL preparatory course. Their TOEFL scores ranged from 457 to 621. All participants signed release forms permitting their test results to be used in this research project in accordance with the ethical standards of the Publication Manuel of the American Psychological Association (1999).

Materials

Accuracy Level Test (ALT)

This test is a published 100 word vocabulary test that has a ten minute time limit, although most students finish before the time limit (Carver, 1987b). Each item consists of a one-word stem and three possible answers, one of which is a synonym of the word in the stem. Items begin at the first grade level and progress to university level according to the Carroll, Davies, and Richman (1971) ranked list. Published test-retest reliability for this test is .86.

In this research the scores derived from the ALT were used to measure written vocabulary knowledge (as opposed to aural vocabulary knowledge). This test is a measure of accuracy level (AL) in rauding theory. With native speaker subjects, it measures accuracy of comprehension, meaning that the higher the score on this test, the more difficult a passage a subject can read. One point was given for each correct answer and subjects were told not to answer if they did not know the word. The K-R21 reliability coefficient for this test was 0.88.

Auding Accuracy Level Test (AALT)

The AALT is a measure of Oral Vocabulary Knowledge. In rauding theory it is a measure of Verbal Level (VL). For non-native speaker subjects, however, it is unlikely that VL can be validly measured since it is supposedly a correlate of crystallized intelligence (Carver, 1999). Tests of crystallized intelligence try to estimate how much an individual knows about the world by measuring vocabulary size, however, there is no logical reason why knowledge of the world and L2 vocabulary size should be related. Thus the AALT should be only considered valid for estimating how many words an individual knows auditorily. Each item consists of a one-word stem presented auditorily at regular intervals and three possible written answers, one of which is a synonym of the word in the stem. Items begin at the first grade level and progress to university level based on the Carroll, Davies, and Richman (1971) ranked list. Both the AALT and the ALT use the same words, a convention continued from all of Carver's prior research dealing with rauding theory. One point was given for each correct answer and subjects were told not to answer if they did not know the word. The K-R21 reliability coefficient for this test was 0.87.

Rate Level Test (RLT)

This is a two-minute rate of reading test. It is a speeded test that reflects how fast a subject can accurately comprehend a simple English passage (Flesch Reading Ease = 91.2, Flesch-Kincaid Grade Level = 2.7). It is a measure of rate level (RL) in rauding theory. The RLT is a modified cloze in which every fourth word has been deleted. Subjects must choose between the correct word and an obviously wrong word, for instance, “The girl went (eat / to) the store.” The task is to mark as quickly as possible the correct words in the text.

Scores were reported by calculating the total number of characters read, including blank spaces. 10 characters, the average number of characters between each item, were subtracted for each incorrect answer. The published test-retest reliability for this test is 0.84. 28 subjects took both Form A and Form B of the test. The correlation between the two forms was 0.78.

Natural Reading Speed Test (NRS)

This test asked subjects to read an easy passage (Flesch Reading Ease = 96.2, Flesch-Kincaid Grade Level = 2.6) for one minute and then mark where they were reading.

Scores were reported by calculating the total number of characters read, including blank spaces. This test is hypothesized to be a measure of RL. 28

subjects took both Form A and Form B of the test. The correlation between the two forms was 0.88.

True False Rate of Comprehension Test (TF)

This test asked students to answer as many true/false questions as they could in a one-minute period. The language and questions were controlled to avoid unfamiliar words or ambiguity concerning the correct answer (Flesch Reading Ease = 85.5, Flesch-Kincaid Grade Level = 2.9). “Red is a color,” for instance, is true, while “A flower is an animal” is not true. This test is hypothesized to be a measure of RL. Scores were reported by calculating the total number of characters read, including blank spaces. Ten characters, the average number of characters in each item, were subtracted for each incorrect answer. A total of 28 subjects took both Form A and Form B of the test. The correlation between the two forms was 0.91.

Pronunciation Level Test (PLT)

A modified form of The Woodcock Word Reading test (form G) was used to measure pronunciation level (PL). In this test, participants pronounced 20 pairs of increasingly difficult words on a tape, for a total of forty items. One rater listened to the tape and judged if the pronunciation was correct. Items pronounced as they would be in Japanese, i.e. /bu:ku/ instead of /buk/, were marked as incorrect. The correlation between each half of the test was 0.75. The Spearman Brown Prophecy Formula calculated the full test reliability as 0.86.

Non-Word Pronunciation Level Test (DK)

A modified form of The Woodcock Non Word Reading test (form G) was used to measure pronunciation level (PL). In this test, participants were pronounced 20 pairs of increasingly longer pseudo-words on a tape, for a total of forty items. One rater listened to the tape and judged if the pronunciation was correct. Items pronounced as they would be in Japanese, i.e., adding unnecessary vowels to final consents (epenthesis), were marked as incorrect. The correlation between each half of the test was 0.78. The Spearman Brown Prophecy Formula calculated the full test reliability as 0.87.

Phonological Processing Skill Test (Pho)

This test used pseudo-word matching to measure phonological processing skill. It is hypothesized to be a measure of decoding knowledge (DK). The task was modeled after the one used in Nassaji and Geva (1999). It consisted of 35 items, each asking whether or not a pair of pseudo-words had the same pronunciation. *Flem* and *Flemb*, for instance, have the same pronunciation, while *Tup* and *Tupe* have different pronunciations. A total of 28 participants took both Form A and Form B of the test. The correlation between the two forms was 0.67.

The Speed of Thinking Test (STT)

The STT has been used in previous research on rauding theory to provide a measure of Cognitive Speed (CS). Based on the Posner Task (Posner & Mitchell, 1967), it requires individuals to decide as fast as possible whether pairs of letters have the same or different names. For example, *Aa* and *bB* have the same names but *bA* and *aB* have different names. The raw score was determined by counting the total number of items that were completed correctly in two minutes. A total of 28 participants took both Form A and Form B of the test. The correlation between the two forms was 0.78.

Test of English as a Foreign Language (TOEFL)

This test is the standard test of English ability used by universities in North America. It is used primarily for admissions purposes and is considered to distinguish well with scores ranging between 450 and 600. It is divided into three sections: listening, structure, and reading comprehension. The published reliabilities and standard errors of measurement are as follows: for the entire test, .95 and 13.9; for the Listening Section, .90 and 2.0; for Structure and Written Expression (Structure), .86 and 2.7; and for Reading Comprehension (Reading), .89 and 2.4 (ETS, 1997). The K-R21 reliability coefficients (derived from the raw TOEFL scores) for this administration (92 students) were: Listening, 0.82; Structure, 0.74; and Reading, 0.77.

Procedure

After obtaining permission from the administration of the Temple University IELP program, data collection took place on three different days over a one month period. In week one, the three TOEFL classroom teachers handed out the permission forms and administered two tests during regular class time. In week two, the students took the Institutional TOEFL at Temple University . The final week the researcher visited the participant’s classrooms and administered the remaining tests. Data collection for the other participants took place in two sessions. In the first session the students completed the permission forms and

all non-TOEFL testing. In the second session they completed a practice TOEFL provided by the Educational Testing Service. This test was an actual TOEFL that had been used between June 1992 and June 1994.

CHAPTER 4

RESULTS

Assumptions

The assumptions of multivariate normality and linearity were evaluated through SPSS and EQS. Several of the variables had values for skew and kurtosis above 1.0. The *z* score values for these (S - 0 / SE = z, where S = obtained skewness value and SE = standard error for skewness) were non-significant at $p = .01$. After replacement of missing values with the mean (see below), values for skew and kurtosis were still non-significant at $p = .01$. Nevertheless, visual examination of histograms showed several of the variables to be peaked and non-normally distributed. Thus, the variables were transformed using a square root transformation.

After transformation there were three univariate outliers and two multivariate outliers among the subjects in the sample. One participant had an unusually low score on PHO (z=-3.71). On examination of her test, this student was found to have left 23 items unanswered. Another participant had an unusually low score on DK (z=-3.62). On examination of the test this individual was found to have heavily relied on Japanese sounding pronunciations for DK

Table 5. Descriptive Statistics

Variable	M	SD	Kurtosis	Skewness	Minimum	Maximum	Valid N
R_RLT	864.77	217.41	1.20	0.89	403	1659	92
R_TF	647.59	154.16	0.17	0.57	350	1033	92
R_NRS	1088.32	348.83	1.28	1.15	616	2247	88
CS_STT	132.50	20.21	-0.20	-0.58	80	169	54
A_AALT	28.70	11.82	0.68	1.06	12	65	92
A_ALT	28.13	12.37	0.64	1.15	13	65	91
A_LT	35.62	14.00	-0.40	0.62	9	69	92
E1	47.49	4.72	0.59	-0.10	35	59	92
E2	48.74	5.72	-0.27	0.24	36	65	92
E3	46.36	6.37	0.60	0.57	33	63	92
P_DK	29.37	5.57	-0.80	-0.51	17	38	68
P_PHO	19.58	4.71	-0.18	-0.33	8	29	71
P_PLT	30.21	3.53	-0.80	-0.36	22	36	68

but not PLT. Both were deleted. Using Mahalanobis distance two multivariate outliers were detected and deleted ($p<.001$). One was found to have very high scores on the ALT and AALT but an unusually low score on the other vocabulary test, LT. This individual was also a univariate outlier, with an unusually low score on the listening section of the TOEFL ($z = -3.96$). The other multivariate outlier had a high score on the AALT but low scores on the ALT and LT.

An examination of scatterplots and p-p plots found the transformed variables to be linearly related and normally distributed. An examination of conditioning indexes and variance proportions found evidence of multicollinearity among the variables. Of the ten roots, seven had conditioning indexes greater than 30. Among these, three had at least two variance proportions greater than .50. The variables with high variance proportions were ALT, AALT, RLT, NRS, and PLT. ALT, RLT, and PLT were deleted from the analysis and the problem was resolved.

Singularity was assumed to be not present because all analyses ran smoothly. Finally, an examination of the scatterplots of predicted values for the variables measuring EL against the residuals for the variables measuring AL, RL and PL showed normality, linearity, and homoscedasticity.

There were 11 missing values for TF; four for NRS, 21 for PHO and 24 for DK. These were replaced by the mean. All analyses were performed with 92 participants unless otherwise stated. For NRS and TF, the missing values were due to one of two reasons, either the student did not write a name on the test or was absent for one of the testing days. PHO and DK had similar problems. Some participants forgot to write a name on their paper and others were absent on the test day. The large number of missing values for these two variables, however, was due to one of the classroom teachers not being able to administer them due to time restrictions. The decision to replace missing values with the mean rather than deleting the variables was made for two reasons. First, replacement of missing values with the mean does not change the mean of the distribution as a whole. It reduces the variance, which probably reduces the correlations between the other variables, making it a conservative method of handling missing data. Second, the variables with the highest number of missing values, those measuring PL, were considered theoretically important enough to leave in the analysis. The raw data and covariance matrix are available in Appendix 2.

Analysis 1

The Hypothesized Model

Using EQS, relationships were examined between General English Ability (EL), a latent factor with three indicators (TOEFL listening, structure, and reading comprehension), Speed of Comprehension (RL) a latent factor with two indicators (TF and NRS), Vocabulary Knowledge (AL) a latent factor with two indicators (AALT and LT), and Pronunciation Level (PL), a latent factor with two indicators (DK, and PHO).

The hypothesized model is presented above in Figure 2. Circles represent latent variables, rectangles represent measured variables. The absence of a line connecting variables implies lack of a hypothesized direct effect. Figure 2 illustrates the hypothesis that General English Ability (EL) is predicted by two correlated main factors, Speed of Comprehension (RL) and Vocabulary Knowledge (AL). Pronunciation Level is hypothesized to have no direct effect on EL, but to correlate with AL. The original rauding model has PL and RL correlate (shown here by a curved line connecting the two factors). However, for non-native speakers this may not be significant.

Model Estimation

Maximum likelihood estimation was employed to estimate the model. The independence model testing the hypothesis that all variables are uncorrelated

was easily rejected, $X^2(36, N = 92) = 398.248$, $p < .05$. The hypothesized model was tested next. Support for the hypothesized model was found in terms of a non-significant chi-square value and high CFI value, $X^2(22, N = 92) = 30.930$, $p > .05$, CFI = .975. The Bentler-Bonett Normed Fit Index and Non-Normed fit indexes were also high, .922 and .960, indicating good model fit.

Effects

Figure 7 below shows the final model and the standardized solutions for the parameter estimates. The squared multiple correlations (SMC) for the measured variables were as follows: TF, .59; NRS, .36; LT, .79; AALT, .62; PHO, .60; and DK, .35. Looking at the measurement model, one point of interest was the relatively low values for NRS. This was no surprise since the re-analysis of Nassaji and Geva's data (above) had suggested that measures of natural reading speed were measuring something other than just rate level. Indeed, in that study, the measure of natural reading rate did not even load on the rate factor. Instead, it loaded on the accuracy factory.

Looking at the structural model, the standardized solution for the path between RL and EL was .41; for AL and EL it was .57. Both of these were significant at $p < .05$. The proportion of variance in EL accounted for by AL, RL and PL ($R = 1\text{-}D^2$) was .78. There was a significant correlation between AL and

RL (.61) as well as between AL and PL (.27). As predicted, the correlation between PL and RL (.27) was not significant at the .05 level.

Analysis 2

A standard multiple regression was performed between total TOEFL scores as the dependent variable (EL) and variables measuring RL, AL and PL as the independent variables. Analysis was performed using SPSS REGRESSION. Data, participants, and evaluation of assumptions are stated above. Table 6 below displays the correlations between the variables. Table 7 shows the unstandardized regression coefficients (*B*), the standardized regression coefficient (Beta), the squared semipartial correlations (SR^2), R^2 and adjusted R^2. *R* for regression was significantly different from zero, $F(6, 85) = 23.31$, $p < .05$. For the three regression coefficients that differed significantly from zero, 95% confidence limits were calculated. The confidence limit for (square root of) TF were .035 to .151, those for (square root of) NRS were .007 to .072, and those for (square root of) LT were .292 to .753.

Only three of the IVs contributed significantly to prediction of TOEFL scores. The squared semipartial correlation for TF was .04, NRS .03, and LT.09. These three IVs contributed another .46 in shared variability. All together, 62% (60% adjusted) of the variation in TOEFL scores was predicted by knowing scores on these three IVs.

Table 6. Correlations between Variables Measuring AL RL, PL and EL

	r_tf	r_nrs	a_lt	a_aalt	p_pho	p_dk	e_ttl
r_tf	---						
r_nrs	0.389*	---					
a_lt	0.490*	0.233*	---				
a_aalt	0.493*	0.188	0.821*	---			
p_pho	0.215*	0.107	0.237	0.133	---		
p_dk	0.158	-0.027	0.155	0.057	0.413*	---	
e_ttl	0.593*	0.407*	0.717*	0.621*	0.154	0.065	---

*$p < .05$

Table 7. Standard Multiple Regression of Variables Measuring PL, RL and AL on Total TOEFL Scores

Factor	Variable	*B*	Beta	SR^2 (unique)
PL	PHO	-.08	-.04	0.00
PL	DK	-.10	-.04	0.00
RL	TF	.10*	.26	0.04
RL	NRS	.04*	.18	0.03
AL	AALT	.01	.01	0.00
AL	LT	.52*	.55	0.09

R^2 = .62* (Unique variability = .16; shared variability .46)
Adjusted R^2 = .60*
R = .79*
*p <.05

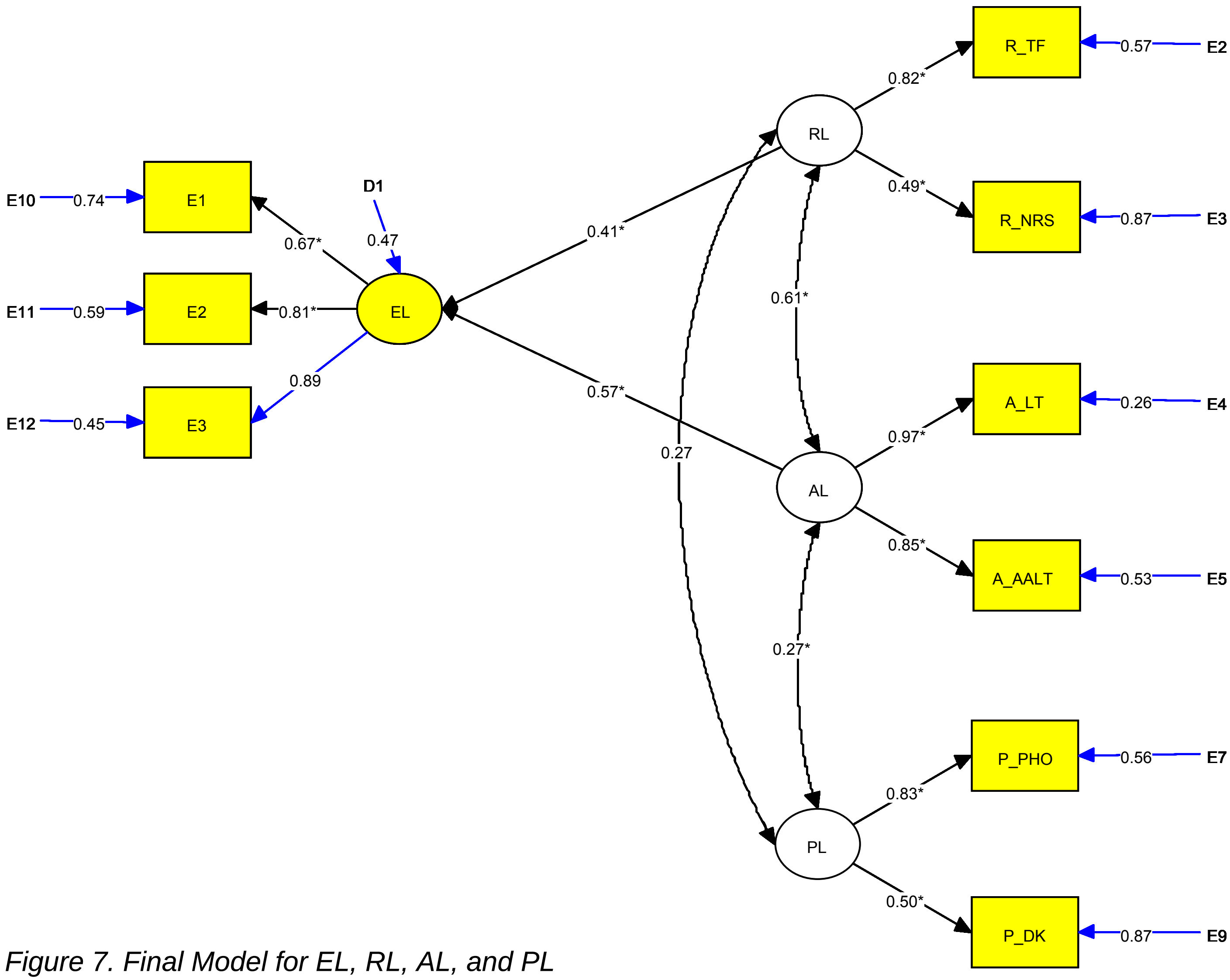

Figure 7. Final Model for EL, RL, AL, and PL

Although the correlation between AALT and TOEFL was .62, AALT did not contribute significantly to regression. Apparently the relationship between TOEFL scores and the AALT was mediated by LT.

Analysis 3

According to rauding theory EL equals (AL + RL) / 2. Using this equation, analysis 3 investigated the validity of this assertion. Data, participants, and assumptions were the same as above except that *z*-scores were used. The *z*-scores for AALT and LT were averaged to create a score for AL. The *z*-scores for NRS and TF were averaged to create a score for RL. Then the variable EL was created following the equation (AL + RL) / 2. The correlation between this derived EL and participants' total TOEFL scores was .78, $p < .05$.

Analysis 4

A pilot study suggested that RL might correlate higher with the listening section of the TOEFL than with other sections. Analysis 4 investigated this assertion. Data, participants, and assumptions were the same as in analysis 3. The correlations between RL and the three sections of the TOEFL were: E1 (listening), .57; E2 (structure), .56; and E3, .49.

Analysis 5

Rauding theory states that RL should correlate with cognitive speed (CS). The reanalysis of Nassaji and Geva's 1999 study (above) found that a measure of CS significantly loaded on the latent factor RL. Analysis 4 investigated whether or not CS correlated with RL. Data, participants and assumptions were the same as in analyses 3 and 4 with the notable exception that only 54 of the participants took the STT. The correlation between RL and the STT (N = 54) was .44, $p < .05$.

CHAPTER 5

DISCUSSION

Of the nine hypotheses presented at the beginning of this study, seven were supported by the analyses, one was inconclusive, and one was rejected. Hypothesis one stated that RL would correlate with EL at an alpha level of 0.05. Analysis 1 found the strength of the path between EL and RL to be .43, which was significant at .05. It should be noted that this was similar to the value found in the SEM reanalysis of Oller's (1972) data above, in which the strength of the path between RL and EL was 0.38. Thus, hypothesis one was supported.

Hypothesis two stated that AL would significantly correlate with EL at an alpha level of .05. Analysis 1 found the strength of the path between EL and AL to be .57, which was significant at .05. It should be noted that this was also similar to the value found in the SEM reanalysis of Oller's (1972) data above, in which the strength of the path between AL and EL was 0.51. Thus, hypothesis two was supported.

Hypothesis three stated that EL as measured by the equation (RL + AL) / 2 would correlate at least 0.75 with a direct measure of EL, the total TOEFL scores. Analysis 3 found a correlation of .78, $p < .05$ between the derived measure of EL and the direct measure of EL. Thus, hypothesis three was supported.

Hypothesis four stated that RL would correlate higher with the listening section of the TOEFL than with the other sections. Analysis 4 found that while RL did correlate highest with the listening section of the TOEFL, it was only one hundredth of a point higher than the correlation with the structure section. Thus, hypothesis four was inconclusive.

Hypothesis five stated that RL would correlate with CS at an alpha level of .05. Analysis five found that RL and CS were correlated at .44 ($p < .05$). Thus, hypothesis five was supported.

Hypothesis six stated that RL would not correlate significantly with PL at an alpha level of .05. Analysis one found that the correlation between PL and RL was .27, which was not significant at the stated alpha level. It should be noted that this was similar to the value found in the SEM reanalysis of Nassaji and Geva's (1999) data, in which the strength of the path between RL and PL was 0.27 and also not significant at the alpha level of .05. Thus, hypothesis six was supported.

Hypothesis seven stated that RL and AL would be correlated at an alpha level of .05. Analysis 1 found the strength of the path between AL and RL to be .61, which was significant at the stated alpha level of .05. It should be noted that this was similar to the values found in the SEM reanalysis of Oller's (1972) data above as well as Nassaji and Geva's (1999) data, in which the strengths of the paths between RL and AL were .44 and .53 respectively. Thus, hypothesis seven was supported.

Hypothesis eight stated that PL would equal AL. Analysis 1 found the strength of the path between AL and PL to be .27. Although this was significant at an alpha level of .05, it was much lower than was expected based on the reanalysis of Nassaji and Geva's study reported above. It was also not close to 1.0, which had been predicted. Thus, hypothesis eight was not supported. This, however, could be due to the fact that a large number of missing values were replaced by the mean for PL, which probably reduced the correlation coefficient.

Hypothesis nine stated that the data would fit the proposed structural model illustrated in Figure 2. Analysis 1 found a non-significant chi-square and a CFI of .98, indicating good model fit. Thus, hypothesis nine was supported.

Conclusions

The data and subsequent analyses presented above tentatively support the applicability of the Rauding Model with non-native speakers of English. This has four main implications that are relevant to theory building and pedagogy in second language acquisition. First, further research is necessary with the rauding model. Second, language proficiency, or performance on tests of language ability, should be defined as EL, or efficiency level. Third, performance differences in tests measuring listening and reading ability may be attributable to RL rather than PL. Finally, in the Japanese educational context, instruction focusing on improving RL may be necessary in the national curriculum.

There are, however, many unanswered questions. Will the strengths of the paths between factors remain consistent among groups of different abilities? Will it be possible to predict scores on different measures of EL such as tests of speaking or writing ability? Not to mention, as in vital with any correlational research, can a standard unit of measurement be created? Rauding theory has used Grade Equivalency Units for native speaker children, but for non-native speakers, a unit of measurement based on the TOEFL or another widely used measure of EL may be more appropriate.

If there is a standard unit of measurement and enough research has been done to determine the relative strengths of association between AL, RL and EL for certain ability levels, then experimental research is necessary. Is it possible, for instance, to increase EL by giving a group instruction aimed at increasing RL or AL? It is widely accepted that increasing vocabulary knowledge, or AL, leads to an increase in a measure of EL such as the TOEFL. This is not true, however, for RL. Indeed, it would not be surprising if many educators questioned the effectiveness of exercises using simple vocabulary to increase comprehension speed in a TOEFL preparatory course. Yet herein lies the most important and exciting implication of rauding theory for second language education in Japan.

RL is a measure of speed of comprehension, meaning that when an individual understands all of the vocabulary and grammar in a text, it measures how fast they can comprehend. Measures of RL use simple language and are administered in only one or two minutes. The measures of RL above that

correlated highest with the TOEFL were a natural reading speed test and a speeded true false test, both using very simple vocabulary. How can such measures explain 40% of the variance in TOEFL scores, a much longer test containing extremely difficult vocabulary?

Theories in second language acquisition such as Anderson's Adaptive Control of Thought (ACT) Model (1985) and McLaughlin's information processing model (1990) have long asserted that there are two types of knowledge. Procedural knowledge (or procedural skill) and declarative knowledge are two common terms used to describe them. Implicit and explicit knowledge have been used to describe these concepts as well. Accuracy level in rauding theory corresponds with declarative knowledge while rate level corresponds with procedural skill. While the ACT model has been described as "enormously complex" (Ellis,1994), the rauding model is simple and elegant. In addition, the rauding model is easily tested (and thus can be easily falsified) and, unlike the ACT model, has clearly defined variables that are easily related to second language instruction and learning. This gives the rauding model predictive power and clear pedagogical implications for educators and researchers in the field of language learning.

Rauding theory can be used to explain the outcome of certain educational programs, such as the one designated by the Japanese Ministry of Education for English education. The English curriculum in Japanese junior and senior high schools focuses on increasing AL, while mostly ignoring RL. Students are

required to memorize an enormous amount of vocabulary and grammatical rules, yet there is no focus on understanding simple English quickly. As a result, students struggle in any situation where they are expected to use English in "real-time," such as in speaking or listening. Could a focus on the development of RL with Japanese school-age learners result in a significant increase in EL? This question is perhaps the most exciting one to be answered in the Japanese educational context.

In conclusion, the results of this study have four main implications. First, the rauding model needs to be investigated further in the context of second language acquisition. Second, language proficiency, or performance on tests of language ability, can be defined as efficiency level. Third, performance differences in language tests purporting to measure reading and listening proficiency may be attributable to the latent factor RL, not PL. Finally, instruction focusing on improving RL should be incorporated into the Japanese national curriculum.

From the viewpoint of theory building, the rauding model improves on prior models of language acquisition. This is because of the salience and predictive power of the factors AL, RL and EL. By defining language proficiency as language efficiency, research and classroom goals become clearer. In addition, the rauding model has clear pedagogical value in the prediction of an individual's language ability and diagnoses of weaknesses in individuals or

language programs. This is especially true in the Japanese educational context, where instruction has traditionally shied away from RL focused instruction.

Limitations

The greatest limitation to this study is the number of missing values distributed among the variables, particularly those associated with measures of PL. Initially, 130 subjects participated in the research, but only 92 completed enough of the tests to be eligible for analysis. While a significant correlation was found between AL and PL (.26), this was much lower than the value found in the reanalysis of Nassaji and Geva’s study (.67). The replacement of missing values with the mean resulted in non-normal distributions, particularly high kurtosis, which then had to be transformed in order to complete the analysis.

The problem with multicollinearity among the measured variables was another weakness. The ALT and AALT used the same items, but were presented differently: either visually or auditorally. This resulted in an extremely high correlation between the two variables and therefore problems with multicollinearity. There were similar problems among DK, PLT and NRS, and RLT. It is uncertain, however, why these variables had multicollinearity problems. Refinement of the instruments used to measure AL, RL and PL is necessary for further research.

Another issue is that one study with only 92 participants cannot be used to generalize to any large population. More research with larger numbers of subjects is necessary in order to stabilize parameter estimates and determine if differences exist in the model among groups of different ability levels.

Finally, correlation does not imply causation, and therefore experimental research is necessary to determine if manipulation of AL and RL will lead to direct changes in EL.

Suggestions for Further Research

There are three suggestions for further research. These can be broken down as ways to: (a) reduce the problem of missing values, (b) improve measurement of AL, and (c) increase the overall efficiency of the testing process.

There were two things that could have been done to reduce the number of missing values. First, all testing should have been done on a single day to avoid missing data. Second, the tests should have been put together in a single package so that participants only would have had to write their name one time. Doing these two things would be likely to reduce the number of missing values in the end and should be considered in any future research.

When measuring AL, the ALT and AALT should use different words to avoid problems with multicollinearity. In addition, a test measuring depth of

vocabulary knowledge (such as John Read's Word Associates Test) could be used to measure AL along with Nation's Levels Test, the ALT and the AALT.

Finally, for research of this size it might be better and more cost effective to create computerized testing programs that would allow researchers to administer all tests in a computer room. While the cost of developing software would be initially expensive, the time and money needed to grade over one thousand paper tests is also substantial. Developing testing software would eliminate scoring and data entry error, as well as permit more accurate measurements of rate. Word recognition speed, for instance, may be the key to explaining why there is such a strong correlation between measures of rate and accuracy. Computerized vocabulary tests have the ability to not only score whether or not an individual knows a word, but also how quickly that individual can identify the meaning of the word.

In conclusion, modifying the tests used in this research and developing software to administer them in a computer room are the next steps that should be taken in the quest to investigate what fundamental factors underlay performance on tests of general language ability.

REFERENCES

American Psychological Association. (1999). *Publication manual of the American psychological association* (4th ed.). Washington D.C.: American Psychological Association.

Anderson, J. (1985). *Cognitive psychology and its implications* (2nd ed.). San Francisco: Freeman.

Anderson, J.R., Lebiere, C., & Lovett, M. (1998). Performance (pp. 57-100). In Anderson, J.R. & Lebiere, C. (Eds.). *The atomic components of thought*. Mabwah, NJ: Lawrence Erlbaum.

Arevart, S., & Nation, P. (1991). Fluency improvement in a second language. *RELC Journal, 22*, 84-94.

Bae, J., & Bachman, L.F. (1998). A latent variable approach to listening and reading: Testing factorial invariance across two groups of children in the Korean/English two-way immersion program. *Language Testing, 15*(3), 380-414.

Buck, G. (1992). Listening comprehension: Construct validity and trait characteristics. *Language Learning 42*(3), 313-357.

Blau, E.K. (1990). The effect of syntax, speed, and pauses on listening comprehension. *TESOL Quarterly, 24*(4), 746-53.

Bradly, L., & Bryant, P.E. (1983). Categorizing sounds and learning to read – a causal connection. *Nature, 301*, 419-421.

Carroll, J.B., Davies, P., & Richman, B. (1971). *Word frequency book*. New York: Houghton Mifflin.

Carver, R. (1977). Toward a theory of reading comprehension and rauding. *Reading Research Quarterly, 13*, 8-63.

Carver, R. (1982). Optimal rate of reading prose. *Reading Research Quarterly, 18*, 56-88.

Carver, R. (1983). Is reading rate constant or flexible? *Reading Research Quarterly, 18*, 190-215.

Carver, R. (1984). Rauding theory predictions of amount comprehended under different purposes and speed reading conditions. *Reading Research Quarterly, 19,* 205-218.

Carver, R. (1985a). How good are some of the world's best readers? *Reading Research Quarterly, 20*, 389-419.

Carver, R. (1985b). Measuring absolute amounts of reading comprehension using the rauding rescaling procedure. *Journal of Reading Behavior, 17*(1), 29-53.

Carver, R. (1987a). *Reading comprehension and rauding theory*. Kansas City, Missouri: Revrac Publications.

Carver, R. (1987b). *Technical manual for the accuracy level test*. Kansas City, Missouri: Revrac Publications.

Carver, R. (1987c). *Technical manual for the rauding efficiency level test*. Kansas City, Missouri: Revrac Publications.

Carver, R. (1990a). *Reading rate: A review of research and theory*. New York: Academic Press.

Carver, R. (1990b). Predicting accuracy of comprehension from the relative difficulty of material. *Learning and Individual Differences, 2*, 405-422.

Carver, R. (1992a). Rauding theory. *Journal of Reading, 36*(2), 84-95.

Carver, R. (1992b). Effect of prediction activities, prior knowledge, and text type upon amount comprehended: using rauding theory to critique schema theory research. *Reading Research Quarterly, 27*(2), 165-174.

Carver, R. (1992c). What do reading standardized reading comprehension tests measure in terms of efficiency, accuracy, and rate. *Reading Research Quarterly, 27(*3), 346-359.

Carver, R. (1992d). The three factors in reading ability: Reanalysis of a study by Cunningham, Stanovich, and Wilson. *Journal of Reading Behavior, 24*(2), 173-190.

Carver, R. (1993). Merging the simple view of reading with rauding theory. *Journal of Reading Behavior, 25*(4), 439-455.

Carver, R. (1997). Reading for one second, one minute, or one year from the perspective of rauding theory. *Scientific Studies of Reading, 1*(1), 3-43.

Carver, R. (2000). *The causes of high and low reading achievement*. New Jersey: Lawrence Erlbaum Associates.

Carver, R., & Darby, C., Jr. (1972). Analysis of the chunked reading test and reading comprehension. *Journal of Reading Behavior, 5*, 282-296.

Cermak, L.S. (1983). Information processing deficits in children with learning disabilities. *Journal of Learning Disabilities, 16*, 599-605.

Cohen, J., McClelland, J., & Dunbar, K. (1990). On the Control of Automatic Processes: A Parallel Distributed Processing Account of the Stroop Effect. *Psychological Review, 97*(3), 332-361.

Cunningham, A.E., Stanovich, K.E., & Wilson, M.R. (1990). Cognitive variation in adult college students differing in reading ability. In T.H. Carr & B.A. Levy (Eds.), *Reading and its development: Component skills approaches* (pp. 129-159). New York: Academic Press.

Daneman, M. (1991). Individual differences in reading skills. In R. Barr, M.L. Kamil, P.B. Mosenthal, & P.D. Pearson (Eds.), *Handbook of reading research* (vol. 2, pp. 512-538). New York: Longman.

Dornyei, Z. (1995). On the teachability of communication strategies. *TESOL Quarterly, 29*,(1), 55-85.

Esky, D. (1970). A new technique for the teaching of reading to advanced students. *TESOL Quarterly, 4*(4), 16-22.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. New York: Academic Press.

Gough, P.B., & Tunmer, W.E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6-10.

Green, J. (1984). Speech preparation process and verbal fluency. *Human Communication Research, 11*(1), 61-84.

Griffiths, R. (1992). Speech rate and listening comprehension: Further evidence of the relationship. *TESOL Quarterly, 26*(2), 385-390.

Hieke, A.E. (1981). Audio-lectal practice and fluency acquisition. *Foreign Language Annals, 14*(3), 189-194.

Hirai, A. (1999). The relationship between listening and reading rates of Japanese EFL learners. *The Modern Language Journal, 83*, 367-384.

Holmes, J. (1954). Factors underlying major reading disabilities at the college level. *Genetic Psychology Monographs, 49*, 3-95.

Hoover, W.A., & Gough, P.B. (1990). The simple view of reading. *Reading and Writing, 2*, 127-160.

Jackson, M.D., & McClelland, J.L. (1979). Processing determinants of reading speed. *Journal of Experimental Psychology: General, 108*(2), 151-181.

Joreskog, K.G. (1993). Testing structural equation models. In K.A. Bollen & J.S. Long (Eds.) *Testing structural equation models* (p. 297). Newbury Park, CA: Sage Publications.

Kagan, J. (1982). Retrieval difficulty in reading disability. *Topics in Learning and Learning Disability, 3,* 75-83.

Laberge, D., & Samuels, S.J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 189-207.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*, 387-417.

Logan, G.D. (1979). On the use of concurrent memory load to measure attention and automaticity. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 189-207.

Kitao, K. (1994). Getting students to read actively. *Doshisha Studies in English, 63,* 49-78.

Macha, D.H. (1979). Reading comprehension of non-native students in English composition at the freshman level. *TESOL Quarterly, 13*(3), 425-427.

MacNamara, J. (1970). Comparative studies of reading and problem solving in two languages. *TESOL Quarterly, 4*(2), 3-12.

Maruyama, G.M. (1997). Basics of structural equation modeling. Thousand Oaks, California: Sage Publications.

McLaughlin, B. (1990). Restructuring. *Applied Linguistics, 11*, 113-28.

Nassaji, H., & Geva, E. (1999). The contribution of phonological and orthographic processing skills to adult ESL reading: Evidence from native speakers of Farsi. *Applied Psycholinguistics, 20*, 241-267.

Nation, I.S.P. (1990). *Teaching & learning vocabulary*. Boston: Heinle & Heinle.

Norris, W.E. (1970). Teaching second language reading at the advanced level: Goals, techniques, and procedures. *TESOL Quarterly, 4*(1), 18-36.

Oller, J.W. (1972). Assessing competence in ESL reading. *TESOL Quarterly, 6*(4), 25-35.

Palmeri, T.J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, memory and cognition, 23*(2), 324-354.

Perfetti, C.A. (1985). *Reading Ability*. New York: Oxford University Press.

Perfetti, C.A. (1989). There are generalized abilities and one of them is reading. In L.B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 307-335). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pierce, M.E. (1973). Sentence level expectancy as an aid to advanced reading. *TESOL Quarterly, 7*(3), 50-58.

Plaister, T. (1968). Reading instruction for college level foreign students. *TESOL Quarterly, 2*(3), 164-168.

Posner, M.I. & Mitchell, R.F. (1967). Chronometric analysis of classification. *Psychological Review, 74*, 392-409.

Posner, M.I., & Snyder, C.R. (1975). Attention and cognitive control. In R.L. Solso (Ed.), *Information processing and cognition* (pp.55-85). Hillsdale, NJ: Earlbaum.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of non-native speaker conversations. *Discourse Processes, 14*, 423-441.

Saville-Troike, M. (1973). Reading and the audio-lingual method. *TESOL Quarterly, 7*(4), 42-52.

Schneider, W., & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1-66.

Schumacker, R.E. & Lomax, R.G. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.

Share, D.L., Jorm, A.F., Maclean, R., & Matthews, R. (1984). Sources of individual differences in reading achievement. *Journal of Educational Psychology, 76*, 1309-1324.

Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127-190.

Stanovich, K.E., Cunningham, A.E., & Freeman, D.J. (1984). Intelligence, cognitive skills, and early reading processes. *Reading Research Quarterly, 29*, 278-303.

Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360-407.

Thompson, M.E. (1985, October). *Dimensions of speed reading: A review of research literature*. Paper presented at the 28th annual meeting of the North Central Reading Association, Ann Arbor, MI.

Tomiyama, M. (1980). Grammatical errors and communication breakdown. *TESOL Quarterly, 14*(1), 71-79.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics, 17*(1), 84-119.

Tunmer, W.E., & Nesdale, A.R. (1985). Phonemic segmentation skill and beginning reading. *Journal of Educational Psychology, 77*, 417-427.

Zhao, Y. (1997). Speech rate and L2 comprehension. *Applied Linguistics, 18(*1), 49-68.

APPENDIX A

Covariance Matrix and Raw Data

Table 8. Covariance Matrix for Variables Purporting to Measure EL, AL, RL and PL

r_rlt	r_tf	r_nrs	a_lt	a_aalt	a_alt	p_pho	p_plt	p_dk	e1	e2	e3
12.981	5.793	9.080	1.404	1.189	1.332	0.108	0.036	-0.053	0.322	0.672	0.561
5.793	8.978	5.736	1.704	1.552	1.540	0.316	0.138	0.216	0.529	0.689	0.684
9.080	5.736	24.186	1.327	0.974	0.992	0.257	0.181	-0.061	0.715	0.706	0.701
1.404	1.704	1.327	1.347	1.001	1.010	0.134	0.079	0.082	0.200	0.281	0.401
1.189	1.552	0.974	1.001	1.105	1.069	0.068	0.039	0.027	0.151	0.213	0.332
1.332	1.540	0.992	1.010	1.069	1.186	0.065	0.029	0.029	0.138	0.221	0.321
0.108	0.316	0.257	0.134	0.068	0.065	0.240	0.051	0.092	0.027	0.024	0.027
0.036	0.138	0.181	0.079	0.039	0.029	0.051	0.078	0.076	0.005	-0.002	0.012
-0.053	0.216	-0.061	0.082	0.027	0.029	0.092	0.076	0.208	0.007	0.013	0.011
0.322	0.529	0.715	0.200	0.151	0.138	0.027	0.005	0.007	0.118	0.081	0.091
0.672	0.689	0.706	0.281	0.213	0.221	0.024	-0.002	0.013	0.081	0.167	0.138
0.561	0.684	0.701	0.401	0.332	0.321	0.027	0.012	0.011	0.091	0.138	0.214

*missing values have been replaced by the mean (*n*=92)

Table 9. Raw Data before Missing Values Replaced By Mean

no	r_rlt	r_tf	r_nrs	r_stt	a_lt	a_aalt	a_alt	p_pho	p_plt	p_dk	e1	e2	e3	e_ttl
1	946	662	1617	129	27	16	21	25	34	35	47	47	48	473
2	649	509	787	*	24	22	21	19	35	26	47	39	36	407
3	792	350	1735	144	30	19	16	20	34	31	46	50	48	480
4	826	589	993	148	57	22	20	22	35	36	50	52	49	503
5	705	689	1048	*	45	24	26	*	*	*	52	54	50	520
6	1196	900	1139	127	65	52	59	*	*	*	51	56	55	536
7	749	*	902	*	23	20	20	21	25	23	49	48	42	463
8	875	736	1117	152	31	18	19	24	31	37	50	50	40	467
9	553	589	906	*	24	16	16	14	34	29	42	42	45	430
10	*	*	696	*	48	31	42	25	36	31	50	54	48	507
11	836	499	902	98	56	41	44	*	*	*	35	51	49	457
12	948	816	1186	*	62	65	65	*	*	*	51	59	63	573
13	789	625	743	*	30	24	23	23	31	35	53	43	46	473
14	1097	625	1098	142	23	14	14	*	*	*	45	42	39	420
15	1228	900	1270	126	69	59	61	*	*	*	51	59	63	573
16	714	726	834	149	40	29	27	*	33	30	50	50	50	500

Table 9 (continued)

no	r_rlt	r_tf	r_nrs	r_stt	a_lt	a_aalt	a_alt	p_pho	p_plt	p_dk	e1	e2	e3	e_ttl
17	843	406	962	95	38	27	28	17	30	21	43	43	47	443
18	1155	424	760	114	13	12	13	15	23	23	46	47	39	440
19	789	412	844	*	27	20	19	24	29	35	42	44	45	437
20	577	412	731	119	28	25	25	20	27	21	43	45	47	450
21	704	*	629	*	20	17	16	26	31	36	43	45	40	427
22	946	499	646	*	44	34	36	14	26	27	49	54	54	523
23	786	775	902	163	31	24	23	24	34	36	47	46	45	460
24	946	*	2194	128	32	23	26	22	35	36	54	55	45	513
25	1166	579	1598	143	28	30	23	24	29	23	51	50	48	497
26	813	543	1021	154	21	14	13	17	27	33	36	47	37	400
27	674	543	939	104	26	39	47	16	25	34	51	50	40	470
28	685	662	743	*	47	39	38	*	*	*	51	58	50	530
29	594	422	696	*	57	40	48	*	*	*	45	43	45	443
30	759	848	953	132	41	43	40	*	*	*	46	56	57	528
31	1372	890	1356	158	50	45	38	16	28	27	50	54	54	527
32	739	765	1410	*	64	42	50	*	*	*	54	54	53	537
33	684	543	671	124	27	28	22	16	27	29	46	50	46	473
34	946	*	1356	*	20	16	20	18	*	*	45	46	42	443
35	664	642	993	130	32	33	37	27	32	32	45	43	49	457
36	875	662	993	154	23	17	15	*	*	*	47	40	44	437
37	734	*	*	*	22	29	23	28	34	37	52	47	46	483
38	587	*	690	*	21	17	16	20	22	19	46	49	45	467
39	1133	816	1917	148	54	43	47	*	*	*	53	45	55	507
40	*	402	939	*	28	25	23	14	32	29	44	44	44	440
41	1659	1033	1543	157	65	54	52	*	*	*	57	57	63	577
42	1093	858	1116	*	31	25	26	25	31	24	50	53	51	513
43	649	454	1066	*	41	35	24	14	30	18	54	42	49	483
44	1248	1033	1545	150	53	40	39	*	*	*	51	59	55	546
45	573	736	1116	*	35	30	23	21	30	34	53	49	49	503
46	1002	966	1356	141	38	47	38	18	*	*	52	51	42	483
47	1272	976	1471	147	53	48	51	*	*	*	48	59	57	545
48	855	519	1741	140	24	17	16	19	30	24	53	48	44	483
49	845	652	975	*	29	22	22	*	*	*	44	48	43	450
50	835	519	1356	144	20	18	18	10	27	26	45	46	43	447
51	887	689	1080	112	20	13	15	18	27	30	48	51	41	467
52	1008	662	787	*	26	23	21	19	31	32	45	48	43	453
53	1198	766	1812	138	36	20	27	21	25	17	48	52	40	467
54	674	615	966	141	45	32	28	22	31	31	54	52	51	523
55	907	716	1617	143	49	27	24	16	31	36	56	51	46	510
56	587	553	1101	157	27	24	22	29	30	32	46	47	48	470
57	*	454	*	107	38	26	*	17	24	22	46	54	50	500
58	1155	509	760	140	42	31	30	20	33	33	49	55	44	493
59	855	679	844	169	20	16	15	8	34	34	44	49	47	467
60	604	519	616	*	16	14	15	17	28	33	44	40	41	417

Table 9 (continued)

no	r_rlt	r_tf	r_nrs	r_stt	a_lt	a_aalt	a_alt	p_pho	p_plt	p_dk	e1	e2	e3	e_ttl
61	826	519	808	129	50	26	29	23	32	36	47	48	49	480
62	1113	699	1171	139	47	41	38	21	31	31	47	54	50	503
63	546	553	743	127	37	27	25	26	33	34	47	53	45	483
64	724	489	787	80	21	21	19	8	26	19	45	49	48	473
65	826	615	1020	136	45	24	20	21	26	27	50	43	41	447
66	403	391	902	*	17	22	19	15	28	26	40	38	36	380
67	918	625	1303	137	36	26	25	25	33	31	46	43	46	450
68	674	553	*	108	26	32	29	12	30	26	44	46	45	450
69	792	486	890	*	16	29	32	14	24	22	44	40	39	410
70	908	625	851	91	34	29	39	23	30	29	45	45	39	430
71	907	579	1116	*	30	23	26	18	31	31	47	52	44	477
72	897	642	939	127	37	28	23	21	30	38	48	55	49	507
73	563	*	*	*	25	22	19	15	25	23	48	48	44	467
74	1216	1033	1342	122	69	61	60	*	*	*	59	65	63	621
75	*	*	1535	*	45	38	34	23	34	31	46	49	50	483
76	664	652	939	*	26	32	35	18	33	33	44	40	36	400
77	812	900	1328	*	37	22	19	*	*	*	56	58	50	547
78	885	*	760	*	25	18	15	21	34	31	45	44	41	433
79	916	716	1188	146	41	25	23	17	32	27	51	54	51	520
80	966	*	1269	*	27	30	26	22	34	33	46	48	39	443
81	813	543	760	*	30	22	25	10	*	*	36	36	34	353
82	965	553	1392	*	27	25	20	27	35	36	41	40	37	393
83	799	625	923	*	37	27	19	22	28	33	45	50	48	477
84	724	662	687	*	32	20	22	23	28	35	43	45	45	443
85	654	454	1236	112	25	22	21	18	*	*	46	44	48	460
86	1272	943	2247	145	53	42	42	*	*	*	59	59	59	590
87	1344	669	1157	*	9	13	14	17	28	21	36	43	33	373
88	789	736	734	96	60	52	53	*	*	*	51	45	47	473
89	695	765	861	*	27	18	18	17	34	28	47	43	40	433
90	875	765	844	119	58	54	33	24	33	26	46	45	46	457
91	875	726	1085	162	45	35	35	26	36	34	41	50	47	460
92	759	569	1171	112	27	18	17	18	25	19	48	50	41	463

APPENDIX B
TESTS

Speeded True-False Test 1 (TF)

T / F	1. Blue is a color. T / F	15	15
T / F	2. The sky is green. T / F	17	32
T / F	3. Soccer is a sport. T / F	18	50
T / F	4. Gold is very cheap. T / F	19	69
T / F	5. People live in rocks. T / F	21	90
T / F	6. Kyoto is a famous city. T / F	23	113
T / F	7. Paper comes from trees. T / F	23	136
T / F	8. People swim in the ocean. T / F	25	161
T / F	9. You can write with a pen. T / F	25	186
T / F	10. Airplanes are very small. T / F	25	211
T / F	11. New York is a city in Japan. T / F	28	239
T / F	12. Trees are made of hard metal. T / F	29	268
T / F	13. People eat breakfast at night. T / F	30	298
T / F	14. Smoking is bad for your health. T / F	31	329
T / F	15. The police are very bad people. T / F	31	360
T / F	16. Most airplanes fly under water. T / F	31	391
T / F	17. Rice is a popular food in Asia. T / F	31	422
T / F	18. A person is bigger than a mouse. T / F	32	454
T / F	19. People listen to music with CDs. T / F	32	486
T / F	20. Japan is a city in North America. T / F	33	519
T / F	21. People play sports in the library. T / F	34	553
T / F	22. Baseball is a popular game in Japan. T / F	36	589
T / F	23. You don’t need money to go shopping. T / F	36	625
T / F	24. Many people sleep during the daytime. T / F	37	662
T / F	25. Walking is faster than driving a car. T / F	37	699
T / F	26. Two minutes is longer than two hours. T / F	37	736
T / F	27. Many people speak English in the world. T / F	39	775
T / F	28. When you are tired, you should not sleep. T / F	41	816
T / F	29. A tomato should be green when you eat it. T / F	42	858
T / F	30. When you are hungry, you want to exercise. T / F	42	900
T / F	31. A small notebook is usually very expensive. T / F	43	943
T / F	32. Many animals live in houses like people do. T / F	43	986
T / F	33. If you study a lot you can go to a good school. T / F	47	1033
T / F	34. If you don't sleep for a long time you will be tired. T / F	53	1086
T / F	35. If you never practice soccer, you will be a good player. T / F	56	1142

Natural Reading Speed Test (NRS)

□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□start□□□□□□□□□□□□□□□□
□□□
□□□□□□□□□□□□□□□□□□□□□□stop□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□

EXAMPLE

Jenny went to the store to buy some food but it was closed.

READ THIS NOW:

Akiko looked at the river. It was moving fast and looked dangerous. “I have to save that poor cat,” she thought. The cat had been in the tree before the river got big. Now it couldn’t get down, and was making crying noises in the tree.

Akiko took off her shoes and put one foot in the water. It was cold and moving very fast. “I have to be careful,” she thought. She walked slowly to the tree. It was not far, maybe only one meter, but the water was moving very fast, so it was not easy to walk. She moved one foot, and then she moved the other foot. Soon she was at the tree. The cat looked at her.

“Hello,” said the cat. “Thank you for coming to help me.”

Akiko stopped. A talking cat! “You can speak?” she asked the cat.

“Yes,” said the cat. “I learned how to speak at school.”

A cat in school! Akiko was very surprised. “You have a good head,” she said.

“No,” the cat said, “I’m not so smart, but I like English.” The cat looked at the river. “By the way, can we leave now? I really don’t like water.”

This is really strange, Akiko thought to herself, but she took the cat in her arms and carried it back to the ground.

“Thank you very much,” it said, “but I am late for a meeting, so I have to go now.” Then the cat ran away quickly.

Akiko went home. “No one will believe me if I say I saw a talking cat,” she thought, so she kept her mouth shut. But, at dinner that night, she asked her father if he believed in talking animals.

“Dad,” she said, “Do you believe that animals can talk?” Her father looked at her strangely. He didn’t say anything for a minute and then he said, “No, animals cannot talk.”

That night, Akiko couldn’t sleep very well. The next day, when she woke up, she had an idea. I can tell Jenny about the talking cat! she thought. After breakfast, she called Jenny on the telephone. “Let’s meet next to the river at lunch time,” she said.

“That’s a good idea,” Jenny answered. “We can have a picnic by the river.”

That day, at 12:30, Jenny and Akiko met by the river. “I saw a strange thing yesterday,” Akiko said. “I came here in the morning to paint a flower, but when I arrived the flower was gone, and the river was so big that even those three trees were in the water!”

Jenny looked at the three trees. They were about one meter away from the river. “The river was very big,” she said.

“Yes, it was,” Akiko said quickly, “and there was a cat in the big tree. It was crying and wanted to return to the ground. The water was all around the tree, so it couldn’t move. I took off my pants and shoes, went in the river, and walked to the tree. Then a very strange thing happened. The cat said hello to me!!!”

Jenny looked at Akiko and laughed. “I don’t believe you,” she said.

“Really, it did!” Akiko said. “It said ‘thank you’ to me. It was a very nice cat, but it ran away after I helped it.”

That is very strange! Jenny said.

Phonological Awareness Test (Pho)

□□
□□□□□□□□□□□□□○□□□□□□□□□□□□○□□□□□□□□□

1. plemb - plem □□ □□
2. nert - nirt □□ □□
3. munk - mank □□ □□
4. wrax - rax □□ □□
5. rept - repped □□ □□
6. nij - nidge □□ □□
7. kight - keet □□ □□
8. naid - nade □□ □□
9. fype - fip □□ □□
10. sert - surt □□ □□
11. pilk - pilque □□ □□
12. darge - darg □□ □□
13. vour - vore □□ □□
14. pern - purn □□ □□
15. datt - dat □□ □□
16. thake - thack □□ □□
17. mought - muft □□ □□
18. koe - kue □□ □□
19. cern - sern □□ □□
20. creat - crite □□ □□
21. thun - thone □□ □□
22. zign - zine □□ □□
23. taum - tume □□ □□
24. wust - wussed □□ □□
25. cirk - cerk □□ □□
26. gnole - nole □□ □□
27. micked - mict □□ □□
28. joat - jote □□ □□
29. voam - vum □□ □□
30. caf - kaf □□ □□
31. dapi - dappy □□ □□
32. jakko - jahcoh □□ □□
33. valli - varri □□ □□
34. youp - yupe □□ □□
35. savi - sabi □□ □□

Decoding Knowledge Test (DK)

ift	raff
bim	nan
un	fay
gat	roo
oss	pog
poe	weat
plip	dud's
shab	whie
vunhip	nigh
bufty	sy
straced	chad
than't	tadding
twem	laip
adjex	gouch
yeng	zirdn't
gaked	knoink
cigbet	mancingful
wrey	bafmotbem
transibsodge	monglustamer
vauge	gnouthe

Auding Accuracy Level Test (AALT)

Due to copyright restrictions, the following test could not be reproduced here

The Auding Accuracy Level Test

Revrac Publications, Kansas City, Missouri.

Nation Levels Test (LT)

Due to copyright restrictions, the following test could not be reproduced here

The Levels Test

Heinle & Heinle, Boston,MA

A copy of the Levels Test can be viewed in Nation (1990), pages 262 to 274.